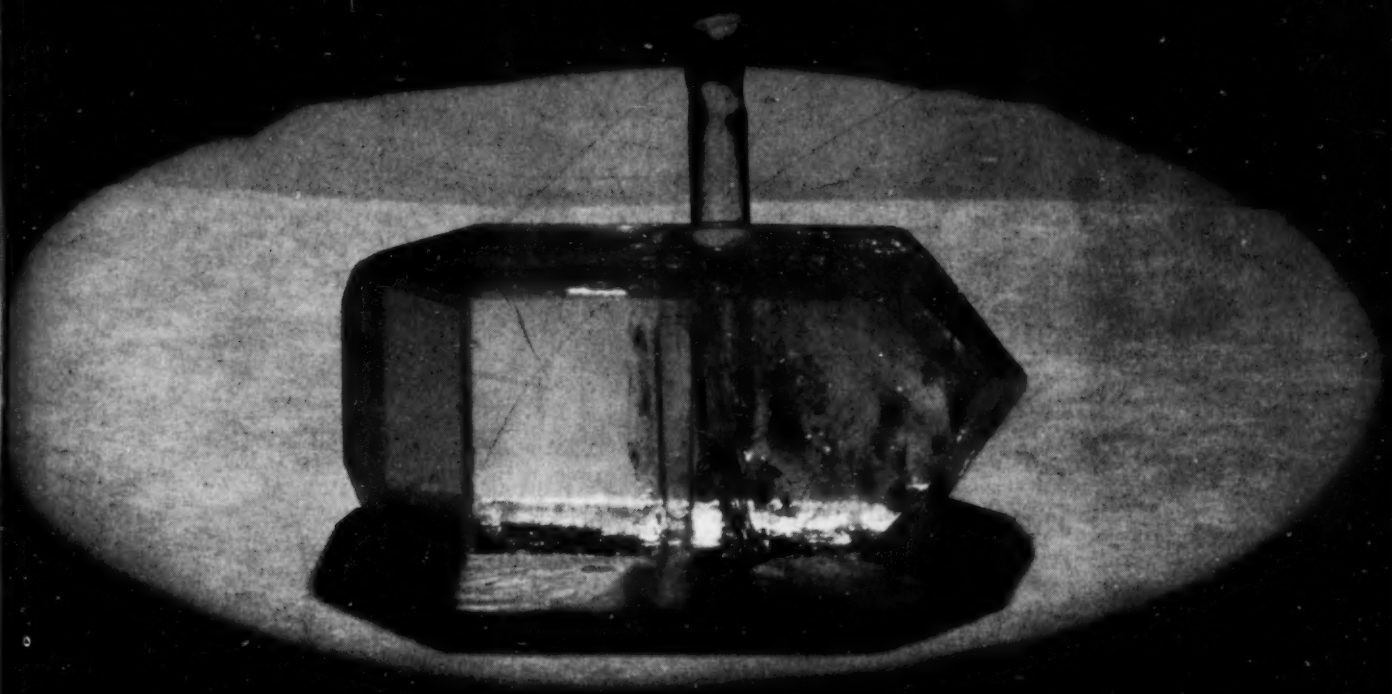
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 22, No. 11 NOVEMBER, 1951

In This Issue

		Page
Necessary Number of Elements in a Directional Ring Aerial	H. Lottrup Knudsen	1299
Response of an Aerodynamic System under External Harmonic Force	H. Serbin	1307
Susceptance of a Thin Iris in Circular Wave Guide with the TM_{01} Mode Incident	Kenneth L. Dunning and Rufus G. Fellers	1316
Predominantly Subharmonic Oscillations	Carl A. Ludeke	1321
Influence of Magnetic Fields upon the Propagation of Electromagnetic Waves in Artificial Dielectrics	E. R. Wicher	1327
General Theory of Symmetric Biconical Antennas	S. A. Schelkunoff	1330
Transit Time Effect in a Cylindrical Diode	Way Dong Woo	1333
Free Longitudinal Vibrations of Rubber and Tissue Strips	Richard W. Lawton and Allen L. King	1340
Application of Nutting's Equation to the Viscoelastic Behavior of Certain Polymeric Systems	Rolf Buchdahl and Lawrence E. Nielsen	1344
Formation of Annealing Twins During Grain Growth	R. L. Fullman and J. C. Fisher	1350
Discrete Energy Levels of a Nonlinear Oscillator	James H. Moran	1355
Concept of Force in Dislocation Theory	M. O. Peach	1359
Atomic Clocks and Frequency Stabilization on Microwave Spectral Lines	C. H. Townes	1365
Nature of the Static and Kinetic Coefficients of Friction	Ernest Rabinowicz	1373
Crystallographic Aspect of Slip in Body-Centered Cubic Single Crystals. I. Theoretical Considerations	A. J. Opinsky and R. Smoluchowski	1380
Erratum: On the Theory of Random Noise. Phenomenological Models. I.	David Middleton	1382
Letters to the Editor:		
Equivalent Temperature of an Electron Beam	M. E. Hines	1385
Comments on "Equivalent Temperature of an Electron Beam"	Philip Parzen	1386
Note on the "Attenuation of Radio Signals Caused by Scattering"	J. B. Smyth and C. P. Hubbard	1386
Reply to J. B. Smyth and C. P. Hubbard	A. H. LaGrone	1386
Distribution of Slip in Metal Crystals	F. C. Frank	1387
Negative Charging of Glass Fibers under Electron Bombardment	John H. L. Watson	1387
Note on the Partial Differential Equations Describing Steady Current Flow in Intrinsic Semiconductors	R. C. Prim	1388
Noble Gas Hot Cathode Diodes with Negative Arc Drop	G. Medicus and G. Wehner	1389
Temperature Function in a Moving Medium	R. H. Ritchie	1389
Books		1390
Other Books Received		1394
Index to Advertisers		xxxii

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Earle C. Gregg, *Assistant Editor*

Associate Editors

Howard H. Aiken
Lyle D. Borst
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
S. A. Goudsmit
G. P. Harnwell
George R. Harrison
F. V. Hunt
Rudolph Kingslake
Paul Kirkpatrick
F. W. Loomis
J. R. Oppenheimer
George B. Pegram
I. I. Rabi
Duane Roller
Harold K. Schilling
Frederick Seitz
Francis G. Slack
John C. Steinberg

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Advertising Manager and Assistant Treasurer*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*
Kathryn Setze *Senior Accountant*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 22, Number 11 November, 1951

The Necessary Number of Elements in a Directional Ring Aerial

H. Lottrup Knudsen
Laboratoriet for Telegrafi og Telefoni, The Royal Technical University of Denmark, Copenhagen, Denmark
(Received October 12, 1950)

An investigation is made concerning the dependence of the array characteristic of a directional ring aerial on its number of elements. It is shown that an odd number of elements is more favorable than an even number in approximating the array characteristic of a similar ring aerial with infinitely many elements. It was previously shown by Page that a similar statement applies to an azimuthly omnidirectional, fading-reducing, concentric ring aerial; this result is contained as a special case in this investigation. The present paper deals especially with the case in which the principal direction of the antenna array is horizontal; the theory for this case is illustrated by a numerical example.

INTRODUCTION

A RING aerial, i.e., an array of s similar and similarly oriented antennas placed equidistantly along a circle, is for sufficiently large s approximately rotationally symmetrical. (For $s=\infty$ the antenna system has complete rotational symmetry.) This approximate rotational symmetry of the ring aerial is used in different ways.

In all the cases that will be dealt with here, the numerical value of the current is assumed to be the same in all elements of the array. Also for the matter of convenience in describing the field, the plane of the circle is assumed horizontal.

Chireix[1] has shown that by phasing the elements progressively so that the phase increases $2\pi n$, where n is an integer, when passing once around the circle, an antenna system with reduced radiation at high elevation angles is obtained. Since the ring aerial obtained in this way is approximately azimuthly omnidirectional for s sufficiently large, it will be useful as a fading reducing antenna for broadcasting purposes. A similar array but with several concentric rings was later independently proposed by Hansen and Woodyard[2,*] and investigated further by Hansen and Hollingsworth.[3,*]

Another type of an omnidirectional antenna with reduced radiation at high elevation angles may be obtained by driving a ring aerial with all currents in the same phase and adding at the center of the ring an antenna carrying a current in phase opposition to the currents in the outer antennas and with an amplitude somewhat less than the sum of the current amplitudes in all of the outer antennas. This array may be considered a special case of one of the arrays investigated by Hansen and Woodyard.[2] The application of this ring aerial as a fading reducing antenna is described by Böhm[4,5] and by Harbich and Hahnemann.[6] The radiation resistance of the above mentioned antenna arrays was calculated by Page.[7] The discrepancy between the array characteristics of ring aerials of the type referred to above with a finite number of elements and the corresponding aerials with infinitely many elements was touched on already by Chireix,[1] by Hansen and Woodyard,[2] and by Hansen and Hollingsworth[3], and was recently investigated in detail by Page.[8] Page showed that in the case of a concentric ring aerial, an odd number of antennas in the ring gives a better approximation of the array characteristic to the array characteristic in the ideal case of infinitely many antennas than an even number of antennas in the ring.

[1] H. Chireix, l'Onde Électr. **15**, 440–456 (1936).
[2] W. W. Hansen and J. R. Woodyard, Proc. Inst. Radio Engrs. **26**, 333–345 (1938).
[*] I am indebted to Dr. J. Epstein of the RCA Laboratories Division, Radio Corporation of America, Princeton, New Jersey, for drawing my attention to these papers.
[3] W. W. Hansen and L. M. Hollingsworth, Proc. Inst. Radio Engrs. **27**, 137–143 (1939).
[4] O. Böhm, Telefunken Zeitung **13**, 60, 21–26 (1932).
[5] O. Böhm, Hochfreq. techn. u. Elektroak. **42**, 137–145 (1933).
[6] H. Harbich and W. Hahnemann, Elektr. Nachr. Technik **9**, 361–376 (1932).
[7] H. Page, Wireless Engineer **25**, 102–109 (1948).
[8] H. Page, Wireless Engineer **25**, 308–315 (1948).

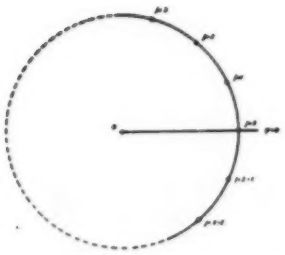


FIG. 1. Ring aerial with s elements.

Stenzel[9] has made an application of the approximate rotational symmetry of a ring aerial quite different from the above mentioned. By giving the currents of the antennas in the ring a phase distribution such that there is constructive interference of the waves emitted in an arbitrary direction in space, hereafter called the principal direction, an array characteristic is obtained having a principal lobe in this direction. Stenzel considers especially the case where the principal direction is horizontal. If the principal direction is rotated in the horizontal plane, the horizontal diagram of the array characteristic will also rotate and, as a consequence of the approximate rotational symmetry of the ring aerial, with its shape almost unchanged. For the ring aerial mentioned here, Stenzel has investigated the discrepancy between the array characteristic in the case of a finite, even number of elements and the characteristic in the case of infinitely many elements, apparently assuming that there is no essential difference between the case of an even number of elements and that of an odd number of elements. This assumption is explicitly stated by Brückmann[10] in his textbook on antennas in the chapter dealing with Stenzel's theory. In the present paper it will be investigated whether the array characteristic of the directional ring aerial with an even number of antennas differs essentially from the characteristic of the ring aerial with an odd number of elements as was the case with the ring aerial investigated by Page.[8] This investigation is based on Stenzel's as well as on Page's theory. For the sake of completeness those parts of Stenzel's and Page's calculations that are used in the present investigation are included in this paper.

THE ARRAY CHARACTERISTIC FOR A RING AERIAL WITH AN ARBITRARY PRINCIPAL DIRECTION

Let s identical and identically oriented antennas be placed equidistantly along a circle with center O and radius ρ as shown in Fig. 1. A polar coordinate system (r, θ, φ) is introduced with its center at O, with its axis $\theta=0$ perpendicular to the plane of the ring aerial (the horizontal plane) and with the plane $\varphi=0$ passing through antenna s. In this coordinate system the azimuth, β_j, of antenna j is expressed by the equation

$$\beta_j = j2\pi/s.$$

The currents, I_j, in the s antennas are assumed to have the same numerical value, but different phases. Using as the time factor $e^{-i\omega t}$, we express I_j as follows

$$I_j = I_0 e^{i\delta_j}.$$

The array characteristic of the ring aerial with the center O as reference point and with the current sI_0 as reference current is then expressed in the following way

$$G=\frac{1}{s}\sum_{j=1}^{s} e^{i[\delta_j - k\rho \sin\theta \cos(\varphi-\beta_j)]},$$

where k denotes the specific transmission coefficient of the medium surrounding the antenna array.

The currents in the antennas are now given such phases, according to Stenzel's[9] proposal, that the waves emitted in an arbitrarily chosen direction (θ_0, φ_0), the principal direction, are in phase. This is obtained by choosing

$$\delta_j = k\rho \sin\theta_0 \cos(\varphi_0-\beta_j).$$

With this choice of the phases δ_j the expression for the array characteristic becomes

$$G=\frac{1}{s}\sum_{j=1}^{s} e^{ik\rho[\sin\theta_0 \cos(\varphi_0-\beta_j)-\sin\theta \cos(\varphi-\beta_j)]}.$$

Following Stenzel we now define an angle ξ by the equation

$$\cos\xi = \frac{\sin\theta\cos\varphi-\sin\theta_0\cos\varphi_0}{[(\sin\theta\cos\varphi-\sin\theta_0\cos\varphi_0)^2+(\sin\theta\sin\varphi-\sin\theta_0\sin\varphi_0)^2]^{\frac{1}{2}}}.$$

Further introducing

$$\rho' = \rho[(\sin\theta\cos\varphi-\sin\theta_0\cos\varphi_0)^2 + (\sin\theta\sin\varphi-\sin\theta_0\sin\varphi_0)^2]^{\frac{1}{2}},$$

we may now express the array characteristic in the following way

$$G=\frac{1}{s}\sum_{j=1}^{s} e^{-ik\rho' \cos(\xi-\beta_j)}.$$

Let us first consider the case where the number of antennas is infinitely large. Introducing in the expression for G

$$s\to\infty,$$

$$\beta_j\to\beta,$$

$$\Delta\beta_j = 2\pi/s\to d\beta,$$

the sum occurring in this expression is converted into

[9] H. Stenzel, Elektr. Nachr. Technik **6**, 165–181 (1929).
[10] H. Brückmann, *Antennen* (Verlag von S. Hirzel, Leipzig, 1939), p. 113.

an integral and we find

$$G=\frac{1}{2\pi}\int_0^{2\pi} e^{-ik\rho'\cos\beta}d\beta=\frac{1}{\pi}\int_0^{\pi}\cos(k\rho'\cos\beta)d\beta=J_0(k\rho'),$$

where $J_\gamma(z)$ denotes the bessel function of order γ and argument z. This expression will be further discussed later and illustrated by diagrams in the special case where the principal direction is horizontal.

We now return to the case where the number of antennas is finite. For further transformation of the above developed expression for the array characteristic we then use

$$\exp(iz\sin\alpha)=J_0(z)+\sum_{n=1}^{\infty}[\exp(in\alpha)+(-1)^n\exp(-in\alpha)]J_n(z).^{11}$$

Setting

$$\alpha=v-\pi/2$$

we find

$$\exp(-iz\cos v)=J_0(z)+\sum_{n=1}^{\infty}(-i)^n[\exp(inv)+\exp(-inv)]J_n(z)$$

$$=J_0(z)+2\sum_{n=1}^{\infty}(-i)^n\cos nvJ_n(z).$$

By using this formula we may express the array characteristic in the following way

$$G=J_0(k\rho')+2\sum_{n=1}^{\infty}(-i)^nJ_n(k\rho')1/s\sum_{j=1}^{s}\cos n(\xi-\beta_j).$$

Recalling the meaning of β_j,

$$\beta_j=j2\pi/s,$$

we may make use of the familiar formula

$$\frac{1}{s}\sum_{j=1}^{s}\cos n\left(\xi-j\frac{2\pi}{s}\right)=\begin{cases}\cos ps\xi \text{ for } n/s=p, \text{ where } p=0,\pm1,\pm2,\cdots,\\ 0 \text{ in any other case.}\end{cases}$$

Substituting in the expression for the array characteristic we obtain

$$G=J_0(k\rho')+2\sum_{p=1}^{\infty}(-i)^{ps}J_{ps}(k\rho')\cos ps\xi.$$

We now have to treat separately the case where the number of antennas is even, and the case where the number is odd.

When s is even:

$$G=J_0(k\rho')+2\sum_{p=1}^{\infty}J_{ps}(k\rho')\cos\left(\frac{\pi}{2}-\xi\right)ps.$$

When s is odd:

$$G=J_0(k\rho')+2\sum_{p=1}^{\infty}J_{2ps}(k\rho')\cos\left(\frac{\pi}{2}-\xi\right)2ps-i2\sum_{p=0}^{\infty}J_{(2p+1)s}(k\rho')\sin\left(\frac{\pi}{2}-\xi\right)(2p+1)s.$$

The first term in the expression for the array characteristic as well in the case of s even as in the case of s odd is seen to be identical with the array characteristic in the case of $s=\infty$. The first term in each of the above expressions may be considered the principal term and the remaining terms correction terms, which express the difference between the array characteristic for the array in question and the array characteristic for the corresponding array with infinitely many elements. Since a bessel function of constant argument decreases rapidly with increasing order, it follows that the array characteristic for an increasing number of antennas, s, converges rapidly to the array characteristic for $s=\infty$ as was expected. As a rule it is desirable to choose s so large that the array characteristic is a good approximation to the array characteristic in the case of $s=\infty$; on the other hand, for practical reasons it is desirable to choose s as small as possible.

The array characteristic in the case of s even is derived by Stenzel approximately in the way described here. The correction terms in the case of s even are seen to be in phase with the principal term. On the other hand, in the case of s odd, there occur correction terms in phase with, as well as correction terms 90° out of phase with, the principal term. Let us for a moment suppose that we have chosen s so large, that the correction terms are small compared with the maximum value of the principal term, i.e., 1. The array characteristic may then sufficiently accurately be expressed by the principal term and the first term in each of the infinite series of correction terms.

When s is even:

$$G=J_0(k\rho')+2J_s(k\rho')\cos(\pi/2-\xi)s,$$

$$|G|=|J_0(k\rho')+2J_s(k\rho')\cos(\pi/2-\xi)s|.$$

When s is odd:

$$G=J_0(k\rho')+2J_{2s}(k\rho')\cos(\pi/2-\xi)2s-i2J_s(k\rho')\sin(\pi/2-\xi)s,$$

$$|G|=([J_0(k\rho')+2J_{2s}(k\rho')\cos(\pi/2-\xi)2s]^2+[2J_s(k\rho')\sin(\pi/2-\xi)s]^2)^{\frac{1}{2}}.$$

Apart from the immediate neighborhood of the zeros

[11] G. N. Watson, *Theory of Bessel Functions* (The Cambridge University Press, Cambridge, England, 1944), p. 22.

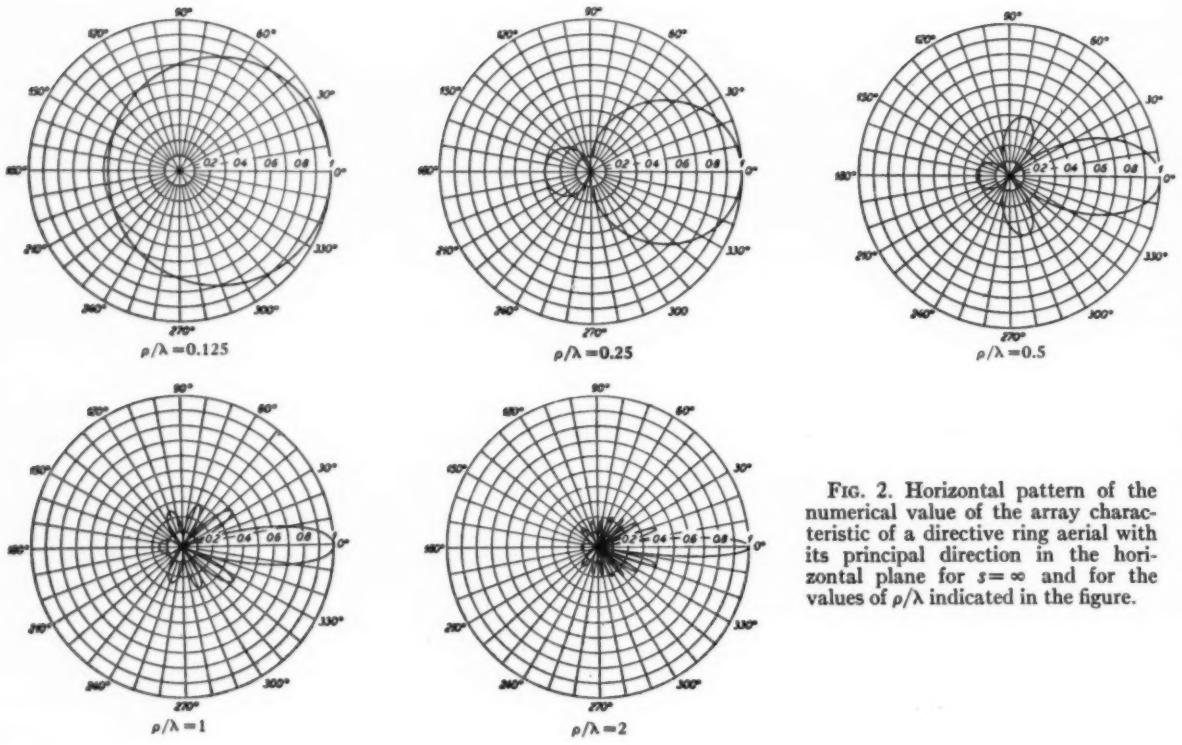


FIG. 2. Horizontal pattern of the numerical value of the array characteristic of a directive ring aerial with its principal direction in the horizontal plane for $s=\infty$ and for the values of ρ/λ indicated in the figure.

of $J_0(k\rho')$ this may be expressed approximately as

$$|G| = |J_0(k\rho')+2J_{2s}(k\rho')\cos(\pi/2-\xi)2s| + (2[J_s(k\rho')\sin(\pi/2-\xi)s]^2)/|J_0(k\rho')|,$$

because a bessel function of constant argument decreases so rapidly with increasing order that the correction term containing the bessel function of order $2s$ may be neglected compared with the correction term containing the bessel function of order s.

Since the correction term 90° out of phase with the principal term is the most important term in the case of s odd, this case will be more favorable than the case of s even as regards the approximation to the array characteristic for $s=\infty$. This will be demonstrated later by a numerical example in the case where the principal direction is horizontal. At first, however, we shall give a short treatment of the case where the principal direction is vertical, thereby establishing the connection between Page's work and the present investigation.

THE PRINCIPAL DIRECTION PERPENDICULAR TO THE PLANE OF THE CIRCLE

For a ring aerial with its principal direction perpendicular to the plane of the circle, i.e., $\theta_0=0$, we find

$$\delta_j=0.$$

The currents in all of the antennas consequently are in phase. By adding to this array an antenna at the center of the circle with a current 180° out of phase with these currents we obtain the fading reducing, concentric ring aerial referred to above.

Setting $\theta_0=0$ in the expressions for ξ and ρ' we get

$$\xi=\varphi,$$
$$\rho'=\rho\sin\theta.$$

Substituting in the expressions for the array characteristic we obtain

when s is even:

$$G=J_0(k\rho\sin\theta)+2\sum_{p=1}^{\infty}J_{ps}(k\rho\sin\theta)\cos\left(\frac{\pi}{2}-\varphi\right)ps;$$

when s is odd:

$$G=J_0(k\rho\sin\theta)+2\sum_{p=1}^{\infty}J_{2ps}(k\rho\sin\theta)\cos\left(\frac{\pi}{2}-\varphi\right)2ps$$
$$-i2\sum_{p=1}^{\infty}J_{(2p+1)s}(k\rho\sin\theta)\sin\left(\frac{\pi}{2}-\theta\right)(2p+1)s.$$

The case of s even was discussed by Chireix;[1] however, besides the principal term of the array characteristic he only gives the first correction term. Page[8] investigated in detail as well the case of s even as the case of s odd and concluded that a much better approximation to the array characteristic in the ideal case of $s=\infty$ is obtained by using an odd number of antennas than by using an even number.

Since the special case mentioned here has been discussed at length in the literature, no further discussion of it will be given here.

THE PRINCIPAL DIRECTION IN THE PLANE OF THE CIRCLE

In this paper we shall especially deal with the case where the principal direction of the antenna array is horizontal, i.e., $\theta_0=\pi/2$. Confining our investigation to the horizontal plane $\theta=\pi/2$ we obtain for this value of θ

$$\xi=\pi/2+(\varphi+\varphi_0)/2),$$
$$\rho'=2\rho\sin((\varphi-\varphi_0)/2).$$

By substituting in the expressions for the array characteristic for a ring aerial with s antennas we get

when s is even:

$$G=J_0(2k\rho\sin((\varphi-\varphi_0)/2))+2\sum_{p=1}^{\infty}J_{ps}(2k\rho\sin((\varphi-\varphi_0)/2))\cos[(\varphi+\varphi_0)/2]ps;$$

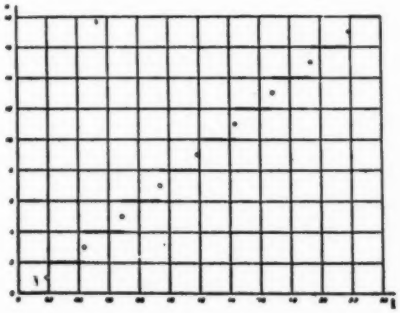


FIG. 3. Number of lobes, n, in the horizontal pattern of the array characteristic of a directive ring aerial with the principal direction in the horizontal plane for $s=\infty$ as a function of ρ/λ.

when s is odd:

$$G=J_0(2k\rho\sin((\varphi-\varphi_0)/2))+2\sum_{p=1}^{\infty}J_{2ps}(2k\rho\sin((\varphi-\varphi_0))/2)\cos[(\varphi+\varphi_0)/2]2ps$$
$$+i2\sum_{p=0}^{\infty}J_{(2p+1)s}(2k\rho\sin((\varphi-\varphi_0)/2))\times\sin[(\varphi+\varphi_0)/2](2p+1)s.$$

The expression for the array characteristic in the case of s even was derived by Stenzel[9] whereas no discussion of the case of s odd seems to have been published. As was already mentioned and as will be further discussed in what follows, the case of s odd differs essentially from that of s even, the first case being more favorable than the second as was the case with the fading reducing, concentric ring aerial investigated by Page.

According to what was shown above, the principal term $J_0(2k\rho\sin((\varphi-\varphi_0)/2))$ is identical with the array characteristic for the corresponding ring aerial with $s=\infty$, so the remaining terms may be considered correction terms. From the expression for the array characteristic in the case of $s=\infty$, for directions in the horizontal plane, we see that it is a function only of the angle v between the direction φ and the principal direction φ_0, $v=\varphi-\varphi_0$. Consequently, when the principal direction is rotated in the horizontal plane an angle $\Delta\varphi_0$ by readjusting the current phases in the way described above, the whole horizontal pattern will be rotated the angle $\Delta\varphi_0$ with its shape unchanged.

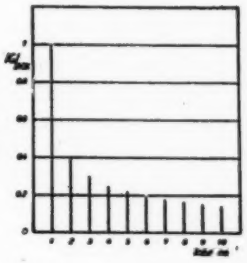


FIG. 4. The maximum value of the fully developed lobes in the horizontal pattern of the array characteristic of a directive ring aerial with the principal direction in the horizontal plane for $s=\infty$.

In Fig. 2 the horizontal pattern of the numerical value of the array characteristic is plotted for $s=\infty$ corresponding to various values of the ratio of the radius ρ of the circle to the wavelength λ. Several of these diagrams are calculated and plotted in the paper by Stenzel[9] referred to above; they are included here for the sake of completeness. From the expression for the array characteristic as well as from the diagrams in Fig. 2 it is seen that the horizontal pattern is symmetrical with respect to $v=0$ and that the number of lobes increases with increasing radius of the circle, whereas the width of the principal lobe decreases. In Fig. 3 the number of lobes, n, is plotted as a function of the ratio of the radius of the circle, ρ, to the wavelength λ. The maximum value of the fully developed lobes in the half-plane $0<v<\pi$ is shown in Fig. 4, where the principal lobe is denoted as number 1.

The width of the principal lobe may be measured by the angle α between the two zero-directions confining this lobe or better by the angle β between the two directions for which the array characteristic assumes $1/\sqrt{2}$ of its maximum value. The angle α is determined as the

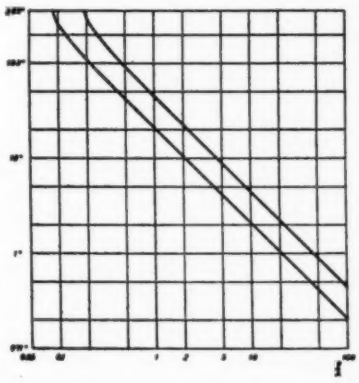


FIG. 5. The angle α between the two zero directions confining the principal lobe in the horizontal pattern of the array characteristic of a directive ring aerial with the principal direction in the horizontal plane for $s=\infty$, and the angle β between the two directions for which the array characteristic is $1/\sqrt{2}$ of its maximum value, plotted as functions of ρ/λ.

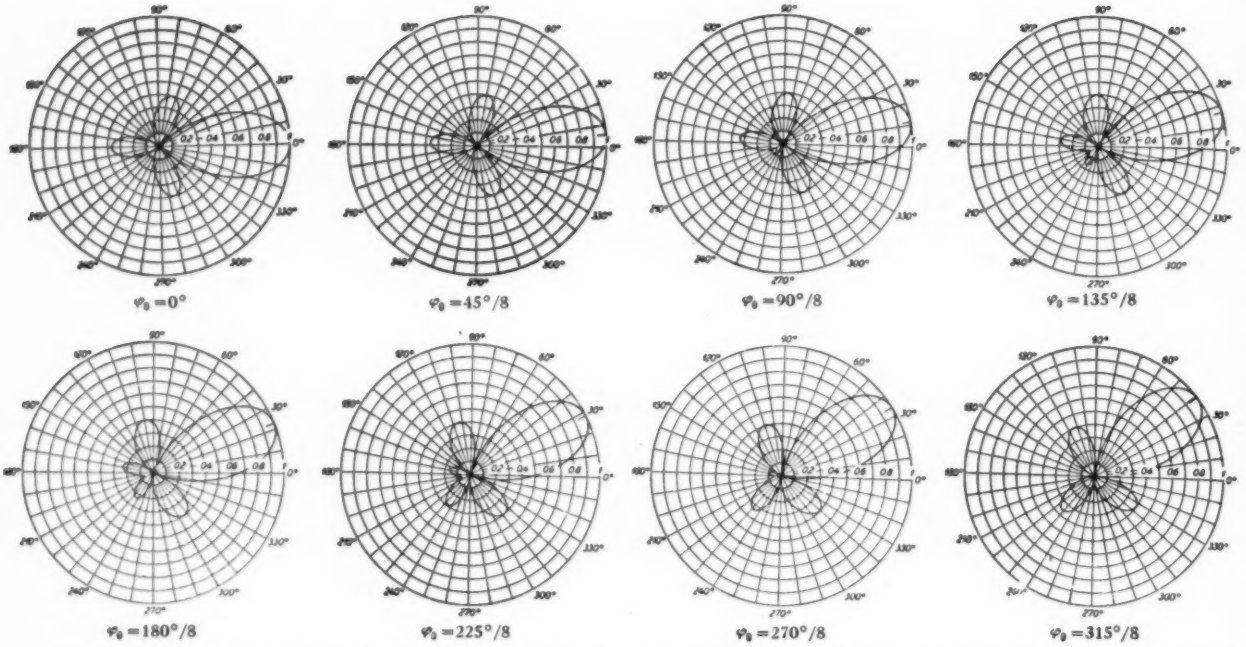


FIG. 6. The horizontal pattern of the array characteristic of a directive ring aerial with the principal direction in the horizontal plane and with $\rho/\lambda=0.5$ and $s=8$ for the values of the principal direction φ_0 indicated in the figure. The position of the antenna system is fixed.

smallest root in the equation

$$J_0(2k\rho\sin(\alpha/4))=0,$$

whereas the angle β is determined as the smallest root in the equation

$$J_0(2k\rho\sin(\beta/4))=1/\sqrt{2}.$$

The angles α and β are plotted in Fig. 5 as functions of ρ/λ.

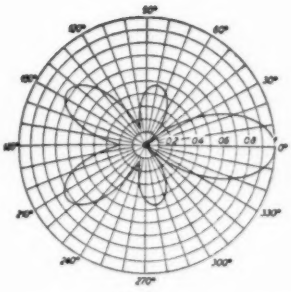
Let us now return to the general case where the number of antennas, s, is finite. Since the array characteristic for horizontal directions, as appears from the expressions derived above, is a function of not only $\varphi-\varphi_0$ but also $\varphi+\varphi_0$, the horizontal pattern will change its shape when the principal direction is rotated by adjusting the current phases in the way prescribed above. From the above derived expressions it is seen that the absolute value of the array characteristic, $|G|$, in the case of s even will have the period in φ_0, $2\pi/s$, whereas in the case of s odd it will have the period π/s. Consequently, we need only investigate $|G|$ for a variation of φ_0 in an interval of the length $2\pi/s$ and π/s for s even and s odd respectively.

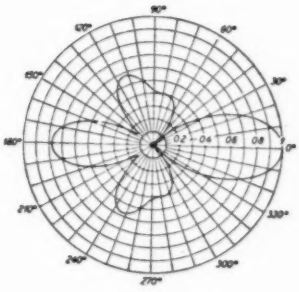
NUMERICAL EXAMPLE

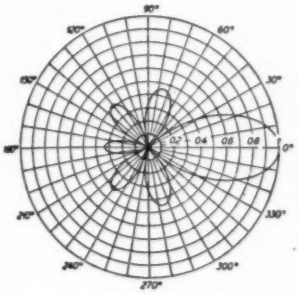
The theory developed above will be illustrated by a numerical example. We consider a ring aerial with the principal direction in the horizontal plane and with $\rho/\lambda=0.5$. For eight antennas in the array, i.e., for $s=8$, and for the principal directions $\varphi_0=0$, $45°/8$, $90°/8$, $\cdots$, $315°/8$ the horizontal pattern of the array characteristic is plotted in Fig. 6. In these diagrams the antenna array is fixed whereas the principal direction rotates as indicated. The figure shows the continuous deformation of the horizontal pattern caused by the rotation of the principal direction. The period is seen to be $360°/8=45°$ in accordance with the theory.

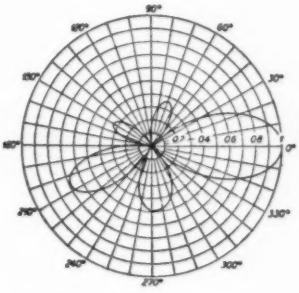
The horizontal pattern of the absolute value of the array characteristic is calculated for $s=5$, 6, 7, 8, and 9. In the case of s even the pattern is calculated for $\varphi_0=0$, $90°/s$, $180°/s$, and $270°/s$, and in the case of s odd for $\varphi_0=0$ and $90°/s$, the pattern for $\varphi_0=180°/s$ in this case being identical with that for $\varphi_0=0$ and the pattern for $\varphi_0=270°/s$ identical with that for $\varphi_0=90°/s$. The results are plotted in Fig. 7 where the horizontal pattern of the absolute value of the array characteristic is plotted as a function of the angle, v, between the direction φ and the principal direction φ_0. Consequently, in the diagrams in Fig. 7 the principal direction is fixed, whereas the array rotates; this way of presenting the results makes it easier to compare the various diagrams.

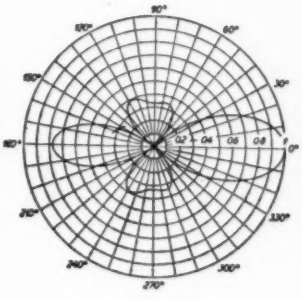
It is impossible to account for the difference between a horizontal pattern corresponding to certain values of s and φ_0 and the horizontal pattern corresponding to $s=\infty$ by a single number. But a consideration of Fig. 7 gives the following result. In the case of only five antennas, the deviations of the horizontal pattern from the pattern corresponding to $s=\infty$ are so large that the minor lobes are of the same order of magnitude as the principal lobe. No improvement is obtained by increasing the number of antennas to six, for the deviations from the ideal pattern are of approximately the same magnitude as in the case of only five antennas.

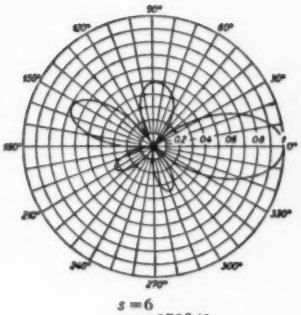

$s=5$
$\varphi_0=0°$

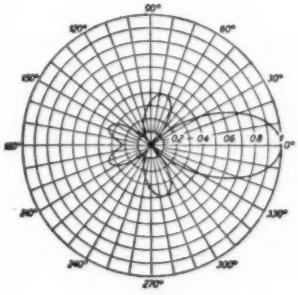

$s=5$
$\varphi_0=90°/5$

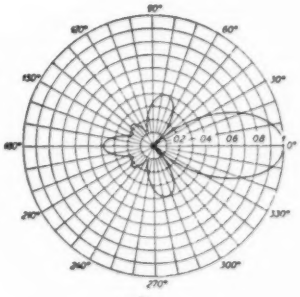

$s=6$
$\varphi_0=0$

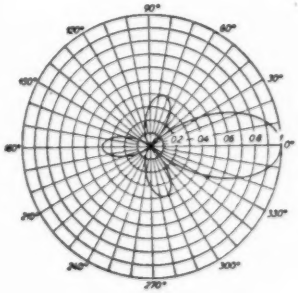

$s=6$
$\varphi_0=90°/6$

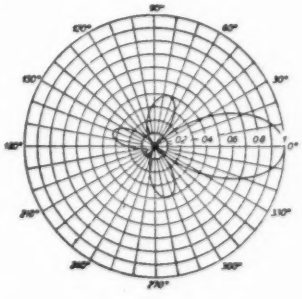

$s=6$
$\varphi_0=180°/6$

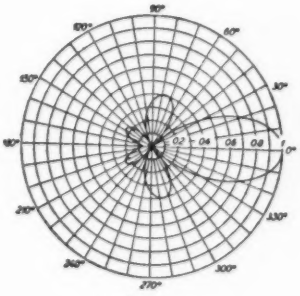

$s=6$
$\varphi_0=270°/6$

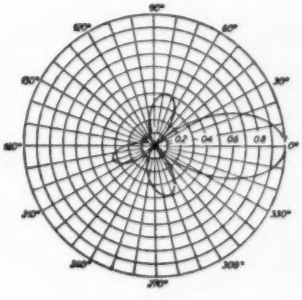

$s=7$
$\varphi_0=0°$

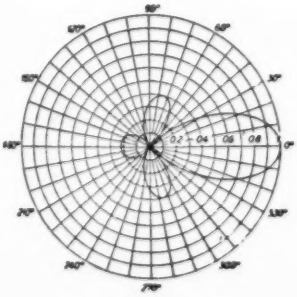

$s=7$
$\varphi_0=90°/7$


$s=8$
$\varphi_0=0°$


$s=8$
$\varphi_0=90°/8$


$s=8$
$\varphi_0=180°/8$


$s=8$
$\varphi_0=270°/8$


$s=9$
$\varphi_0=0°$

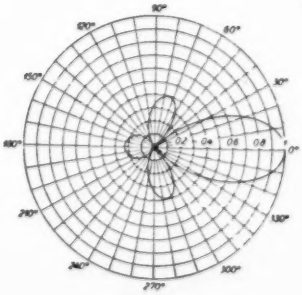

$s=9$
$\varphi_0=90°/9$

FIG. 7. The horizontal pattern of the array characteristic of a directive ring aerial with the principal direction in the horizontal plane and with $\rho/\lambda=0.5$ for the values of the number of antennas, s, and the principal direction, φ_0, as indicated in the figure. The position of the principal direction is fixed.

When the number of antennas is increased from six to seven the situation is improved essentially, for the minor lobes now are essentially smaller than the principal lobe. In the backward direction, however, the horizontal pattern diverges considerably from the ideal pattern. The horizontal patterns corresponding to eight antennas do not approximate the ideal pattern better than do the patterns corresponding to seven antennas. However, an improvement is obtained again when the number of antennas is increased from eight to nine. The deviations of the horizontal patterns corresponding to nine antennas from the ideal pattern are of the same order of magnitude as the accuracy with which the diagrams are plotted. A similar result is found in the case of ten antennas; but the patterns obtained in this case are not plotted in the figure. For a still larger number of antennas the horizontal patterns practically coincide with the pattern corresponding to infinitely many antennas. The cases considered here are seen to group so that the cases $s=5$ and $s=6$ approximate the ideal pattern with the same accuracy and that the same statement applies to the cases $s=7$ and $s=8$ and to the cases $s=9$ and $s=10$, whereas the accuracy changes considerably when we pass from one group to the next one.

The numerical example thus confirms that a ring aerial of the type considered here and with an odd number of elements has a marked advantage over a similar ring aerial with an even number of elements.

Stenzel[9] has shown that the side lobes may be considerably reduced by adding to the ring aerial in question one or more ring aerials concentric with the first one. It would be interesting to make an investigation concerning the influence of the number of elements on the horizontal pattern of this more complex antenna system similar to what has been made here for the simple ring aerial. But this investigation is reserved for another paper.

CONCLUSION

The general expression is derived for the array characteristic of a ring aerial with an arbitrarily chosen principal direction and with a finite number of elements. This expression shows that no uniform improvement of the approximation of the array characteristic to the ideal characteristic is obtained when the number of elements is increased, but that ring aerials with an odd number of elements are more favorable in this respect than aerials with an even number. This was shown by Page[8] for the special case where the principal direction is perpendicular to the plane of the circle. The case where the principal direction lies in the horizontal plane is especially investigated here. A numerical example confirms the above statement concerning the superiority of an odd number of antennas over an even number.

This investigation was carried out at the Laboratoriet for Telegrafi og Telefoni at the Royal Technical University of Denmark, Copenhagen, with a grant from Det teknisk-videnskabelige Forskningsråd. I thank Professor J. Oskar Nielsen, of the Laboratoriet for Telegrafi og Telefoni, for permission to publish this paper. Dr. C. Schulman, of the RCA Laboratories Division, Radio Corporation of America, Princeton, New Jersey, kindly corrected the English translation.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

The Response of an Aerodynamic System under External Harmonic Force*

H. SERBIN
Purdue University, School of Aeronautics, Lafayette, Indiana
(Received December 4, 1950)

In Part I, the theory of nonconservative linear systems of a finite number of degrees-of-freedom is presented. An expression is derived for the response of the system to external harmonic forces in Sec. 3. The effect of small variations in parameters is obtained in Sec. 4. Section 5 is a discussion of the influence additional degree-of-freedom. Part II is concerned with numerical procedure for the calculation of characteristic roots and vectors. The results are applied to a numerical example in Part III.

PART I. THEORY

1. Introduction

IT is a matter of common knowledge among flutter specialists that the airplane is amply endowed with degrees-of-freedom. This considerable freedom of motion manifests itself in many problems, a number of which are included in the expression *aeroelasticity*. On the one hand, at the low frequency range one has the classical airplane dynamics, whereas at the other extreme there are high frequency oscillations confined to local structure such as in the case of a vibrating skin panel. The analysis of any particular problem requires, however, the consideration of a number of degrees-of-freedom; and the appropriate selection of these is one of the crucial points of the analysis.

It is by no means clear that the best answer to the selection problem would be to consider simultaneously a large number of degrees-of-freedom. This procedure may suffer from the disadvantage that the physical interpretation of the computation is obscured.

To some extent one has a criterion for selection of degrees-of-freedom in a conventional, small deflection, vibration problem. For instance, a fixed aileron trim tab of natural frequency 2000 cpm would probably have a small effect on wing frequencies below 1000 cpm. Natural frequencies of the components form an index as to the importance of the relative motion of these components in the composite system.

In the general case in which the three types of forces, inertia, elastic, and aerodynamic, act on the system, the situation is considerably complicated, particularly by the fact that the aerodynamic system of forces is not conservative. The usual vibration theory of small oscillations no longer holds and the concept of a *natural frequency* is not significant. Certain developments, however, can be carried over from the conservative case so that it is possible to extend methods[1] originated to analyze the natural frequencies of composite structures.

The central object in the following investigation is the *response function*, namely, the response of the aerodynamic system under external, harmonic loads. It is expected that this function will be of considerable utility in analyzing the incipient flutter of certain types of composite systems. By using the fourier decomposition theorem, it will be possible to carry over the results into the study of aircraft under the influence of external transient loads of general types, a subject of increasing importance for large aircraft.[2]

The point of view adopted here is that the analytic treatment of the problem achieves the simplest and most basic form when one starts from the conception of virtual work. The element of virtual work dW, due to moving a set of generalized forces $Q_1, Q_2, \cdots, Q_n$ through a virtual displacement $dq_1, dq_2, \cdots, dq_n$ is

$$dW = Q_1 dq_1 + Q_2 dq_2 + \cdots + Q_n dq_n. \qquad (1.1)$$

This equation, in fact, defines the generalized forces Q_i. The utility of the virtual work concept derives from the fact that dW is an invariant. The classical theory of vibrations under conservative forces is primarily one of coordinate transformations, and it is reasonable to expect that the invariance of dW under coordinate transformations will furnish the starting point for a theory of flutter generalizing the conventional vibration theory.

The right side of Eq. (1.1) can be regarded as the inner product of two vectors $(Q_1, \cdots, Q_n)$ and $(dq_1, \cdots, dq_n)$. The inner product of two vectors in two or three dimensions is proportional to the cosine of the included angle, hence, is zero if the vectors are orthogonal. In general, for any number of degrees of freedom, two vectors with zero inner product are *orthogonal*. The two vectors need not be of the same type as, for instance, is the case in Eq. (1.1).

The present treatment makes use of the flutter type of aerodynamic coefficients by considering the behavior of the system under a sinusoidal forcing function. The motion of the system is separated into two parts, one a rigid body motion and the other an elastic deformation which is capable of description in terms of the characteristic modes and frequencies.

The literature dealing with the analysis of flutter is extensive and there are a number of points in which the theory below makes contact with existing developments. The following analysis may then be considered partly expository. The explicit construction of the response

* Prepared under sponsorship of the Air Materiel Command, Dynamics Branch, Contract 33(038) 48-3402-E (1948).
[1] H. Serbin, J. Inst. Aeronaut. Sci. **12**, No. 1, 108–112 (1945).
[2] A. E. Russell, J. Roy. Aeronaut Soc. **51**, 145–170 (1947).

function, the perturbation theory,[3] and the addition of a degree-of-freedom are believed to be new developments.

2. Equations of Motion

Consider a point P which is constrained to move along a straight line so that its coordinate x relative to a fixed reference point varies sinusoidally with time according to

$$x=x_0\sin(\omega t+\delta),$$

and suppose further that an oscillating force

$$F=F_0\sin(\omega t+\epsilon)$$

is applied to the point P. Then the work performed by F in the time interval dt is given by

$$Fdx=F_0x_0\omega\sin(\omega t+\epsilon)\cos(\omega t+\delta)dt,$$

and the work over a complete cycle is

$$\int_{t=0}^{2\pi/\omega}Fdx=-\pi F_0x_0\sin(\epsilon-\delta)$$

$$=-\pi\mathfrak{I}(F\bar{x}).$$

The symbol $\mathfrak{I}$ means "imaginary part of." F and x in the last expression represent

$$F=F_0e^{i\epsilon},$$

$$x=x_0e^{i\delta},$$

and the bar over the x indicates the complex conjugate of this quantity.

In the general case in which a mechanical system is completely specified by n independent coordinates $x_1, x_2, \cdots, x_n$ and is subjected to sinusoidally oscillating forces of a given frequency, the work per oscillation performed by these loads when moved through a displacement $y_1, y_2, \cdots, y_n$ is

$$W=-\pi\mathfrak{I}(\textstyle\sum_r F_r\bar{y}_r), \tag{2.1}$$

where $F_1, F_2, \cdots, F_n$ represent the generalized forces. In the case of small oscillations about a position of equilibrium, the forces F_r are linear functions of the coordinates x and can be expressed in the form

$$F_r=\textstyle\sum_s A_{rs}x_s,\quad r=1, 2, \cdots, n,$$

so that the work (2.3) can be expressed by

$$W=-\pi\mathfrak{I}(\textstyle\sum_{r,s}A_{rs}\bar{y}_rx_s). \tag{2.2}$$

If the system is in equilibrium, it follows from D'Alembert's principle of virtual work that

$$W=0\quad\text{for arbitrary } y, \tag{2.3}$$

which is equivalent to the system of equations

$$\textstyle\sum_s A_{rs}x_s=0,\quad r, s=1, 2, \cdots, n. \tag{2.4}$$

The use of the $\sum$-sign to indicate summation will henceforth be dropped, and the summation convention that a repeated index will indicate a sum with respect to that index will be adopted. Thus, the preceding equation will now be written

$$A_{rs}x_s=0,\quad r, s=1, 2, \cdots, n,$$

where the repetition of the subscript s indicates that the sum on the index s over its range (in this case 1 to n) is intended.

Now consider an aerodynamic system undergoing small sinusoidal oscillations. In general, there are three types of forces (action) on the system, namely, the inertia, aerodynamic, and elastic, each of which contributes to the coefficients of Eq. (2.4). For the present purpose it is convenient, however, to break these coefficients into two parts

$$A_{rs}=\omega^2a_{rs}-b_{rs},$$

where ω^2a_{rs} = virtual work coefficient caused by inertia plus aerodynamic forces and $-b_{rs}$ = virtual work coefficient caused by elastic forces.

Equation (2.4) now takes the form

$$\omega^2a_{rs}x_s-b_{rs}x_s=0,\quad r=1, 2, \cdots, n. \tag{2.5}$$

Dividing through by ω^2 and setting

$$\lambda=\omega^{-2}, \tag{2.6}$$

we obtain for Eq. (2.5)

$$a_{rs}x_s-\lambda b_{rs}x_s=0,\quad r=1, 2, \cdots, n, \tag{2.7}$$

where a_{rs} and b_{rs} are independent of λ.

In practice the aerodynamic parts of the coefficient a_{rs} are based on section data for airfoils oscillating at frequency ω in an air-stream of velocity v and are, therefore, functions of the parameter $k=\omega b/v$, in which b is a reference length. These functions are complex valued in general where the phase displacement relative to the coordinates results from the wake and virtual mass effect.[4]

The case of a forced oscillation under the sinusoidal external forces $F_1, F_2, \cdots, F_n$ leads to the equation,

$$\omega^2a_{rs}x_s-b_{rs}x_s+F_r=0,\quad r=1, 2, \cdots, n,$$

which upon division by ω^2 and the substitution

$$w_r=-\omega^{-2}F_r, \tag{2.8}$$

becomes

$$a_{rs}x_s-\lambda b_{rs}x_s=w_r,\quad r=1, 2, \cdots, n. \tag{2.9}$$

Equation (2.7) is the *homogeneous* case. It is solvable for $x_1, x_2, \cdots, x_n$ with the x's not all zero if and only if the characteristic determinant

$$|A-\lambda B|\equiv|a_{rs}-\lambda b_{rs}|,\quad A=(a_{rs}),\ B=(b_{rs}), \tag{2.10}$$

[3] The author has recently learned of the existence of a similar work by van de Vooren, *A Method to Determine the Change in Flutter Speed Due to Small Changes in the Mechanical System* (National Aeronautical Research Institute, Amsterdam), Report V. 1366.

[4] R. D. Kármán and Sears, J. Inst. Aeronaut. Sci. 5, 379 (1938).

is zero. The roots λ of this equation are the *characteristic numbers* of the mechanical system and can be shown to be independent of the choice of coordinates $x_1, x_2, \cdots, x_n$. The solution $x_1=e_1, x_2=e_2, \cdots, x_n=e_n$ of Eq. (2.7) corresponding to a particular characteristic root λ_ν is called a *direct characteristic mode* and satisfies the equation,

$$a_{rs}e_s-\lambda_\nu b_{rs}e_s=0. \tag{2.11}$$

The flutter problem consists of finding those values of $v/\omega b$ and the corresponding λ for which there is at least one positive characteristic number. The critical $v/\omega b$ and the corresponding λ define the airspeed and frequency of oscillation at (marginal) flutter. The corresponding characteristics vector e defines the critical mode.

Along with Eq. (2.7) consider the adjoint system,

$$y_r a_{rs}-\lambda y_r b_{rs}=0, \quad s=1, 2, \cdots, n, \tag{2.12}$$

which has the solution $y=f$ for $\lambda=\lambda_\nu$:

$$f_r a_{rs}-\lambda_\nu f_r b_{rs}=0. \tag{2.13}$$

The vector f is called a *conjugate characteristic mode.*

If the system of forces is conservative, then it follows that $a_{rs}=a_{sr}$, $b_{rs}=b_{sr}$, and the distinction between direct and conjugate modes disappears. However, aerodynamic forces are usually nonconservative and the theory of linear equations must be extended to cover this case.

3. The Harmonic Response Function

The following developments will be primarily concerned with the *harmonic response function* which is the solution of Eq. (2.9) regarded as a function of frequency ω. This solution is expressed simply in terms of the characteristic numbers and modes:

$$x=\sum_{j=1}^{m}\frac{(f^k, w)}{\lambda_k-\lambda}e^k+g, \tag{3.1}$$

where $m=$Rank (b_{rs}), $\lambda_k=k$th characteristic number, e^k, $f^k=$corresponding direct and conjugate modes, $g=$free body response, and $(f^k, w)=f_1{}^k w_1+f_2{}^k w_2+\cdots+f_n{}^k w_n$.

Equation (3.1) states that the response of the system to a set of harmonic forces is the superposition of free body and elastic response, the latter being proportional to the direct characteristic modes each appearing with a coefficient proportional to the work done by the external force *displaced over the conjugate mode.* The quantity $\lambda_k-\lambda$ appearing in the denominator of the coefficient of e^k is similar to the frequency term in the usual theory of vibration.

The response of the system to a set of forces $Q_r(t)$ varying with time may be found using Eq. (3.1). It is necessary only to separate the forces $Q_r(t)$ into harmonic components, apply Eq. (3.1) to each component, and then to synthesize the resulting harmonic motions.

4. Perturbation Theory

The effect of small changes in the coefficients a_{rs} and b_{rs} will now be considered. Regarding the changes as proportional to a parameter ϵ, replace a_{rs} and b_{rs} by

$$a_{rs}\rightarrow a_{rs}+\epsilon a_{rs}', \tag{4.1a}$$

$$b_{rs}\rightarrow b_{rs}+\epsilon b_{rs}'. \tag{4.1b}$$

The corresponding characteristic numbers and vectors will be transformed as

$$\lambda_\nu\rightarrow\lambda_\nu+\epsilon\lambda_\nu'+\epsilon^2(\cdots), \tag{4.2}$$

$$e^\nu\rightarrow e^\nu+\epsilon e^{\nu\prime}+\epsilon^2(\cdots), \tag{4.3a}$$

$$f^\nu\rightarrow f^\nu+\epsilon f^{\nu\prime}+\epsilon^2(\cdots). \tag{4.3b}$$

Then Eq. (2.11) takes the form

$$(a_{rs}+\epsilon a_{rs}')(e_s{}^\nu+\epsilon e_s{}^{\nu\prime}+\cdots)-(\lambda_\nu+\epsilon\lambda_\nu'+\cdots)(b_{rs}+\epsilon b_{rs}')(e^\nu+\epsilon e^{\nu\prime}+\cdots)=0.$$

Expanding the left side into a power series in ϵ and equating corresponding coefficients on both sides of the equation, we obtain

$$a_{rs}e_s{}^\nu-\lambda_\nu b_{rs}e_s{}^\nu=0, \tag{4.4}$$

$$(a_{rs}e_s{}^{\nu\prime}-\lambda_\nu b_{rs}e_s{}^{\nu\prime})+(a_{rs}'e_s{}^\nu-\lambda_\nu b_{rs}'e_s{}^\nu-\lambda_\nu' b_{rs}e_s{}^\nu)=0, \tag{4.5}$$

etc.

Equation (4.4) furnishes no new information, since it is a repetition of the definition of e^ν. Equation (4.5) is solvable for $e_s{}^{\nu\prime}$ according to Eq. (3.1) if†

$$(f^\nu, a_{rs}'e_s{}^\nu-\lambda_\nu b_{rs}'e_s{}^\nu-\lambda_\nu' b_{rs}e_s{}^\nu)=0,$$

that is, if

$$\lambda_\nu'=f^\nu A'e^\nu-\lambda_\nu f^\nu B'e^\nu, \tag{4.6}$$

where $A'=(a_{rs}')$ and $B'=(b_{rs}')$.‡

An equation similar to Eq. (4.5) can be established for $f^{\nu\prime}$, which would also lead to Eq. (4.6). The vector $f^{\nu\prime}$ can be solved by applying Eq. (3.1) to the adjoints of (a_{rs}) and (b_{rs}).

In the particular case in which the matrix B is unchanged $(B'=0)$, then we have

$$\lambda_\nu'=f^\nu A'e^\nu. \tag{4.7}$$

In words, the change in the system of aerodynamic and inertia forces a_{rs} will result in a change of the virtual work performed by the forces corresponding to the mode e^ν when moved through the virtual displacement f^ν. This change in work will be equal to the change in λ_ν.

The terms in Eqs. (4.2) and (4.3) of the seond and higher orders in ϵ can also be obtained. However, it is possible that the linearized correction term corre-

† We are using the notation for the "inner product"

$$(y, x)=y_1x_1+y_2x_2+\cdots+y_nx_n.$$

If the comma is omitted, read the symbol as the matrix product of the one-column matrix y by the one-rowed matrix x.

‡ Equation (4.6) is in matrix notation.

sponding only to the first power in ϵ will be adequate in practical cases to represent the change.

5. Additional Degrees-of-Freedom

One of the most difficult problems of flutter practice is to assess the effects of an additional degree-of-freedom. Normally, this is handled by a new calculation which simultaneously considers the additional degree together with the original degrees. The considerable increase in numerical labor and complication is familiar to all workers in the field.

The use of the response function brings a considerable simplification to certain problems of this type. Consider the case in which one starts with an aerodynamic system in n degrees of freedom and then adds another degree. The following assumptions will be made.

(*A*) *The basic n degree-of-freedom case is completely solved, that is, the characteristic numbers and vectors are known for all k.*

(*B*) *The additional degree-of-freedom does not introduce additional aerodynamic, inertia, or elastic couplings into the basic degree-of-freedom case, that is, the coefficients* a_{rs} *and* b_{rs} *are unchanged for* $r, s=1, 2, \cdots, n$.

(*C*) *There are no elastic couplings between the additional degree and the original n, that is,*

$$b_{r,n+1}=b_{n+1,r}=0 \quad (r=1, 2, \cdots, n). \tag{5.1}$$

(*D*) *There are no free modes in the system,* $|B|\neq 0$.

Write the system (2.7) in the form

$$a_{rs}x_s-\lambda b_{rs}x_s=-a_{r,n+1}x_{n+1} \quad (r, s=1, \cdots, n), \tag{5.2}$$

$$a_{n+1,s}x_s+(a_{n+1,n+1}-\lambda b_{n+1,n+1})x_{n+1}=P \quad (r, s=1, \cdots, n). \tag{5.3}$$

The solution of Eq. (5.2) can be written on the basis of Eq. (3.1) in the form

$$x_r=-[e_r{}^{\nu}(f_s{}^{\nu}, a_{s,n+1})/(\lambda_{\nu}-\lambda)]\cdot x_{n+1}, \quad r=1, 2, \cdots, n. \tag{5.4}$$

When this is substituted in Eq. (5.3), one obtains the following relation between the external force P and the deflection x_{n+1}:

$$P/x_{n+1}=-[(a_{n+1,r}e_r{}^{\nu})(f_s{}^{\nu}, a_{s,n+1})/(\lambda_{\nu}-\lambda)]+(a_{n+1,n+1}-\lambda b_{n+1,n+1}). \tag{5.5}$$

Under the circumstance that flutter exists, the left side is zero. Therefore, flutter can be detected by plotting the right side in the complex plane as a function of λ and $k=\omega b/v$ and determining the values of these parameters for which the expression is zero.

The right side of Eq. (5.5) can be regarded as the spring constant for the complete system excited in the $(n+1)$st degree of freedom by the external force $Pe^{i\omega t}$. The spring constant is seen to be the sum of several terms of which the last is the "spring constant" of only the $(n+1)$st degree (the other held fixed) and the preceding terms are the coupling effects between the different modes in the n degree-of-freedom case and the $(n+1)$st degree.

Equation (5.5) leads to a simple criterion to insure the absence of flutter. The first term on the right side is a sum of terms of the form $c/(\lambda_{\nu}-\lambda)$, where c is a complex number independent of λ. When this is plotted in the complex plane as a function of the real parameter λ, one obtains a circle passing through the origin of diameter equal to $|c/(\text{Imag }\lambda_{\nu})|$, where the symbol $||$ means that the modulus is to be taken. Therefore, the first term on the right side of Eq. (5.5) represents a complex number falling within a circle of diameter

$$D=\sum_{\nu}|a_{n+1,r}e_r{}^{\nu}|\cdot|f_s{}^{\nu}a_{s,n+1}|/|\text{Imag }\lambda_{\nu}|. \tag{5.6}$$

Therefore, if

$$|a_{n+1,n+1}-\lambda b_{n+1,n+1}|>D, \tag{5.7}$$

flutter cannot exist (at the corresponding value of $v/\omega b$). The expression inside the brackets on the left side is represented by a straight line in the complex plane, and the modulus is always greater than the length of the perpendicular from the origin to the line. In the particular case in which $b_{n+1,n+1}$ is real, this length is $a_{n+1,n+1}$, so that flutter will be avoided if

$$|a_{n+1,n+1}|>D. \tag{5.8}$$

In the more general case in which $b_{n+1,n+1}$ is complex, write $a_{n+1,n+1}$ and $b_{n+1,n+1}$ in the complex form:

$$a_{n+1,n+1}=a_1+ia_2,$$

$$b_{n+1,n+1}=b_1+ib_2,$$

where i is the imaginary unit. Then Eq. (5.8) should be replaced by

$$|a_1b_2-a_2b_1|/(b_1{}^2+b_2{}^2)^{-1}>D. \tag{5.9}$$

The diameters of the circles on the right side of Eq. (5.6) are finite numbers in general because the λ_{ν} are generally complex. In the case where λ_{ν} is real, there will be a flutter condition in the basic n degree-of-freedom case.

Another example in which a degree-of-freedom is added, where, however, assumption (*B*) is not valid, is the case in which the mass m is attached to the system. More generally, suppose that the mass is attached *elastically* to the system at the point P and let the frequency of the mass on its suspension be ω_0. Make the assumption:

(*E*) *There are no aerodynamic forces introduced by the additional degree-of-freedom.*

Now separate the basic system and the sprung mass at the point P. Let x^1 represent the coordinate of the spring in the direction of expansion, and let F be an oscillating force applied to the sprung mass at P. Let x represent the deflection of the basic system measured in the opposite direction from x^1, and let there be applied to the system at P the sinusoidal force F but oppositely

directed. Then there are suitable constants (real) such that

$$x=\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_nx_n, \tag{5.10}$$

where $x_1, x_2, \cdots, x_n$ can be found from Eq. (3.1). On the other hand, looking at the sprung mass, we have

$$F=x^1m\omega_0^2/(1-\omega_0^2/\omega^2). \tag{5.11}$$

Since the consistency of the structure requires that

$$x+x^1=0,$$

it follows that

$$x+F(1-\omega_0^2/\omega^2)/m\omega_0^2=0.$$

The force F can be conveniently chosen as a unit force, so that the equation becomes

$$x+(1-\omega_0^2/\omega^2)/m\omega_0^2=0. \tag{5.12}$$

In particular, if ω_0 is infinite corresponding to a rigid attachment, then it follows that

$$x-1/m\omega^2=0. \tag{5.13}$$

Equations (5.12) and (5.13) can be used to design mass balances to maintain flutter at predetermined speeds.

Consider the elastic attachment of two masses m_1 and m_2 with natural frequencies ω_{01} and ω_{02} at points P_1 and P_2, respectively. Let $x_{rs}(r, s=1, 2)$ be the deflection of the structure at point P_r due to unit oscillating load at P_s. The relation between the force F_r and the deflection x_n^1 of the sprung mass m_r is in accordance with Eq. (5.11):

$$F=x_r^1m_r\omega_{0r}^2(1-\omega_{0r}^2/\omega^2). \tag{5.14}$$

The deflection at P_r due to the reversed forces F_r applied to the structure is

$$x_{r1}F_1+x_{r2}F_2. \tag{5.15}$$

The consistency of deformation requires again that

$$x_{r1}F_1+x_{r2}F_2+x_r^1=0,$$

whence

$$x_{r1}F_1+x_{r2}F_2+(F_r/m_r\omega_r^2)(1-\omega_{0r}^2/\omega^2)=0 \qquad (r=1, 2). \tag{5.16}$$

The values of $v/\omega b$ and ω for which the determinant

$$\begin{vmatrix} x_{11}+\dfrac{1}{m_1\omega_1^2}\left(1-\dfrac{\omega_{01}^2}{\omega^2}\right) & x_{12} \\ x_{21} & x_{22}+\dfrac{1}{m_2\omega_2^2}\left(1-\dfrac{\omega_{02}^2}{\omega^2}\right) \end{vmatrix}=0$$

establish the flutter point of the composite system.

PART II. METHODS OF CALCULATION

1. Introduction

The purpose of Part II is to show how the relations developed are to be translated into numerical procedure. In Part III, an application to a specific example is made.

It is considered desirable to reduce as many as possible of the numerical operations required by the theory to matrix operations, more specifically to operations with real matrices. A matrix with complex elements can be written in the form $A+iB$ with A and B real matrices. If to each such matrix one associates the following one of double the number of columns and double the number of rows

$$\begin{pmatrix} A & -B \\ B & A \end{pmatrix} \tag{1.1}$$

or, instead,

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix}, \tag{1.2}$$

then this correspondence will be preserved under multiplication, addition, and subtraction. Therefore, all such operations can be performed on the complex matrices using either one of the equivalent real forms (1.1) or (1.2).

Let

$$A=(a_{rs}), \quad B=(b_{rs}),$$
$$A+B=(c_{rs}),$$
$$A\cdot B=(d_{rs}).$$

A numerical check of these operations is

$$\sum_s a_{rs}+\sum_s b_{rs}=\sum_s c_{rs},$$
$$\sum_s a_{rs}(\sum_t b_{st})=\sum_t c_{rt}.$$

2. The Direct and Conjugate Modes

Consider the equation

$$Dx-\lambda x=0, \tag{2.1}$$

where D is a square matrix of order m and x a column matrix of order m. One practice for finding the direct characteristic mode x is to form the sequence of vectors $x^1, x^2, x^3, \cdots$, where the first vector x^1 is arbitrarily chosen and the succeeding vectors obtained by iteration:

$$Dx^n-x^{n+1}=0, \quad n=1, 2, \cdots. \tag{2.2}$$

If the sequence of vectors so obtained converges (except for a multiplicative constant) to a vector e^1, then e^1 is a direct mode. The ratio of the components of x^{n+1} to the corresponding components of x^n will then converge to the corresponding characteristic value λ_1. Similarly, the conjugate mode y which satisfies the equation

$$yD-\lambda_1y=0, \tag{2.3}$$

where y is the representation of the vector as a single-row matrix, can be found by forming the sequence y^n iteratively according to

$$y^nD-y^{n+1}=0. \tag{2.4}$$

If the iterations (2.2) and (2.4) would converge for *every* choice of the initial vector x^1 and y^1, then this would be true if the initial vectors x^1 were chosen as the columns of D and the vectors y^1 the rows of D. Hence, in the case of convergence, the direct mode e^1 is the common limiting vector of the columns of Dr and the conjugate mode f^1 is the common limiting vector of the rows of Dr. Hence, it follows that

$$Dr \rightarrow \begin{bmatrix} e_1^1 \\ \cdot \\ \cdot \\ \cdot \\ e_n^1 \end{bmatrix} (f_1^1, \cdots, f_n^1), \tag{2.5}$$

where convergence is to be understood in the sense that two vectors differing only by a scalar multiple are to be considered identical.

Equation (2.5) suggests that the direct and conjugate characteristic modes be obtained by forming a high power of D, say Dr, and obtaining the common column vector and the common row vector. When normalized, these will be the vectors e^1 and f^1, respectively.

In the case $B=U$, the normalization condition (A8) (see Appendix) is

$$f_1e_1+f_2e_2+\cdots+f_ne_n=1;$$

that is, the sum of the terms on the main diagonal of the product matrix

$$\begin{bmatrix} e_1^1 \\ \cdot \\ \cdot \\ \cdot \\ e_1^1 \end{bmatrix} (f_1^1 \cdots f_n^1)$$

is unity. Hence, it follows that

$$\begin{bmatrix} e_1^1 \\ \cdot \\ \cdot \\ \cdot \\ e_n^1 \end{bmatrix} (f_1^1 \cdots f_n^1) = \frac{Dr}{Tr(Dr)}, \tag{2.6}$$

where $Tr(Dr)$, called the "trace of D^n," represents the sum of elements on the main diagonal.

The considerations above are based on the assumption that the iteration process *does* converge. It is known that convergence will take place if there is only one characteristic value of largest modulus. On the other hand, there are important cases where this requirement would not be satisfied. For instance, consider the case where D is a real matrix with complex characteristic values. These must occur in pairs of conjugate complex numbers; hence, there will be at least two of largest moduli. Still another case where convergence can obviously not occur is where D represents a rotation.

The matrix D in flutter depends on $v/\omega b$; and it seems likely, particularly where a large number of degrees-of-freedom is considered, that for some value of the speed-frequency $v/\omega b$ the moduli of two characteristic values will coincide. If $v/\omega b$ is chosen close to such an exceptional value, then convergence may be difficult to obtain.

3. Characteristic Values

When the direct and conjugate modes e^1 and f^1 are known, the characteristic value λ_1 can be obtained by applying Eq. (A8):

$$\lambda_1 = f^1Ae^1. \tag{3.1}$$

4. The Response Function

The response function, Eq. (3.1), Part I, for an arbitrary vector w can be obtained using the forcing vector w and the conjugate modes f.

5. Flutter in $(n+1)$ Degrees-of-Freedom

The condition for flutter in $(n+1)$ degrees-of-freedom is obtained by setting the right side of Eq. (5.5), Part I, equal to zero. The terms of the form $-\alpha_\nu/(\lambda_\nu-\lambda)$ can be transformed by writing

$$\lambda_\nu = a_\nu + ib_\nu,$$

$$\xi = -(a_\nu-\lambda)/b_\nu \quad (b_\nu \neq 0),$$

where a_ν, b_ν, and λ are real, so that

$$-\alpha_\nu(\lambda_\nu-\lambda)^{-1} = (\alpha_\nu/b_\nu)(\xi-i)^{-1}.$$

The values of $1/(\xi-i)$ lie on a circle of unit diameter passing through the origin.

The term $(a_{n+1,\,n+1}-\lambda b_{n+1,\,n+1})$, λ real, is a straight line in the complex plane. Hence, the quantity P/x_{n+1} is readily constructed as the locus of sums of vectors traversing either circles or a straight line. In the exceptional case where b_ν is zero, the circle degenerates into a straight line.

For the particular $v/\omega b$ for which the curve P/x_{n+1} passes through the origin, flutter will exist and for that value of λ at which P/x_{n+1} is zero. For other values of $v/\omega b$, the zeros of Eq. (5.5), Part I, are complex. This suggest that one consider P/x_{n+1} as a function $f(\lambda)$ of λ with λ complex.

Since $f(\lambda)$ is a rational function of λ, then $f(\lambda)$ is analytic. Consider $f(\lambda)$ for λ near λ_0 and expand $f(\lambda)$ into a Taylor's series:

$$f(\lambda)-f(\lambda_0) = c_k(\lambda-\lambda_0)^k + c_{k+1}(\lambda-\lambda_0)^{k+1} + \cdots, \; c_k \neq 0.$$

Hence for sufficiently small $|\lambda-\lambda_0|$, one obtains

$$f(\lambda)-f(\lambda_0) \cong c_k(\lambda-\lambda_0)^k.$$

As λ traverses a small circle in the λ-plane around $\lambda=\lambda_0$ in the *counterclockwise* direction, $f(\lambda)$ will move around $f(\lambda_0)$ *counterclockwise* exactly k times. (Except for a finite number of points, k will equal unity.) One expresses this by saying the $f(\lambda)$ effects a mapping of λ into the $f(\lambda)$ plane which preserves orientation. Therefore, by considering the curve $f(\lambda)$, λ real, and the relation of the curve to the origin, the orientation property will show on which side of the curve $f(\lambda)$ the λ's with positive imaginary part will be transformed. Thus, in the illustration of Fig. 1, the zero of $f(\lambda)$ is in the square

in the λ-plane which borders on the line segment 12; hence, it has positive imaginary part.

Another useful property of $f(\lambda)$ deriving from its analyticity is that $df/d\lambda$ at $\lambda=\lambda_0$ is independent of the manner in which λ approaches λ_0. This means that the transformation $\lambda\rightarrow f(\lambda)$ is locally isotropic and carries small squares into small squares in the $f(\lambda)$-plane. In the case shown in Fig. 1, the magnification can be obtained by comparing the lengths of the segments 12 in the two planes and then used to approximate the complex λ which transforms into the point $f(\lambda)=0$. In this way the complex zeros of $f(\lambda)$ with small imaginary parts can be evaluated.

PART III. EXAMPLE

1. Data and Calculations

The simple example discussed in the present part is the two-dimensional, three-degree-of-freedom case discussed by Theodorsen and Garrick.[5] In the notation of Smilg and Wasserman,[6] one obtains

$$\mu=1/\kappa=4.00, \qquad r_\beta^2=0.0012,$$
$$a=-0.4, \qquad (\omega_h/\omega_\alpha)^2=\tfrac{1}{16},$$
$$x_\alpha=0.2, \qquad (\omega_\beta/\omega_h)^2=\tfrac{3}{2},$$
$$r_\alpha^2=0.25, \qquad 2b=12 \text{ ft},$$
$$c=0.6, \qquad \omega_\alpha=90 \text{ radians/sec},$$
$$x_\beta=0.$$

The first part of the analysis below deals with the bending-torsion subcase, and the results are applied to the bending-torsion-aileron case. The following coordinates are used:

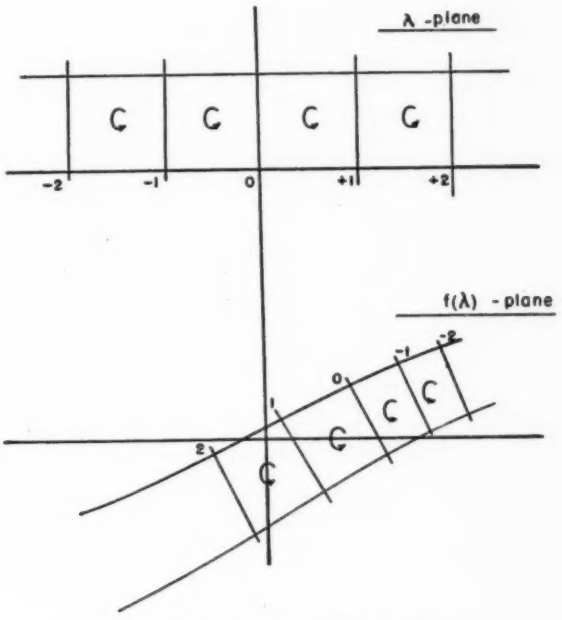


FIG. 1. The transformation $\lambda\rightarrow f(\lambda)$.

x_1 = deflection at the elastic axis,

x_2 = angle of inclination of the airfoil,

x_3 = aileron angle.

The calculations below are made for the parameter $v/\omega b=0.83$. When the conjugate degrees of freedom are chosen the same as the direct degrees (i.e., x_1, x_2, x_3), the aerodynamic-plus-inertia matrix A is

$$A=\begin{pmatrix} 4.8538 & -0.8838i & 0.3323 & -1.5065i & -0.40325-0.18305i \\ 1.2146 & +0.0884i & 1.3718 & -0.6827i & -0.21653-0.19088i \\ 0.022290-0.005618i & & 0.018219-0.037681i & & -0.00715-0.01806i \end{pmatrix}.$$

The elastic matrix B is

$$B=\begin{pmatrix} 0.250 & 0 & 0 \\ 0 & 1.000 & 0 \\ 0 & 0 & 0.00045 \end{pmatrix}.$$

In the two-degree-of-freedom subcase (x_1, x_2), the conjugate degrees of freedom are now chosen as the generalized forces:

$$X_1=0.250x_1,$$
$$X_2=1.000x_2;$$

and the aerodynamic-plus-inertia matrix now is

$$\begin{pmatrix} 19.4152 & -3.5352i & 1.3292 & -6.0260i & -1.61300 & -0.73220i \\ 1.2146 & +0.0884i & 1.3718 & -0.6827i & -0.21653 & -0.19088i \\ 0.022290-0.005618i & & 0.018219-0.037681i & & -0.007147-0.018058i & \end{pmatrix}$$

which will henceforth be denoted by A. The elastic matrix takes the form

$$B=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.00045 \end{pmatrix}.$$

In the two-degree-of-freedom subcase, we obtain the first characteristic mode by forming a high power of

$$A_1=\begin{pmatrix} 19.4152-3.5352i & 1.3292-6.0260i \\ 1.2146+0.0884i & 1.3718-0.6827i \end{pmatrix},$$

$$A_1^8=\begin{pmatrix} 0.01696-0.97933i & -0.29964-0.12990i \\ 0.01686-0.06221i & -0.01703-0.01311i \end{pmatrix}\times 20^8.$$

[5] Theodorsen and Garrick, *Mechanism of Flutter, a Theoretical and Experimental Investigation of the Flutter Problem*, NACA TR 685 (1940).

[6] Smilg and Wasserman, *Three Dimensional Flutter Analysis*, AAF TR 4798.

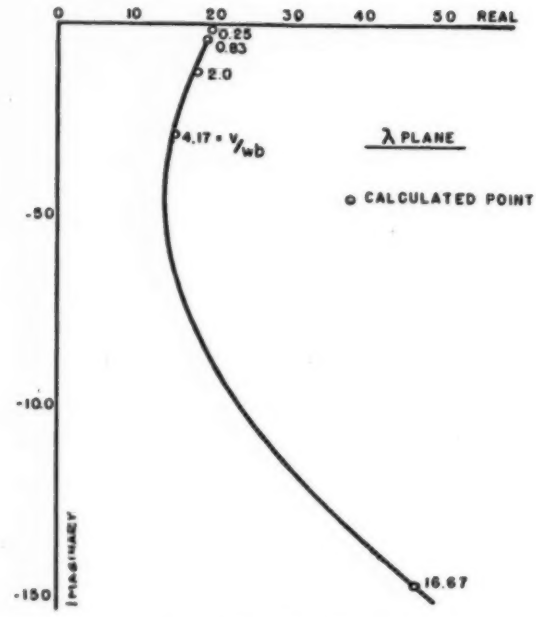


FIG. 2(a). Characteristic root λ_1.

The common column vector $V_x=(V_{x1}, V_{x2})$ is proportional to e^1:

$$V_x=\begin{pmatrix}1\\0.0638+0.0161i\end{pmatrix}.$$

The common row vector $V_y=(V_{y1}, V_{y2})$ is proportional to f^1:

$$V_y=(1,\quad 0.1272-0.3083i),$$

$$\lambda_1=V_yAV_x/(V_y, V_x)$$

$$=20.24\angle -12°\,15'/1.013\angle -1°,$$

$$=19.98\angle -11°\,15'=19.60-3.90i,$$

$$P_1=\frac{1}{Tr(A^8)}\cdot A^8=\begin{pmatrix}0.9868+0.0172i & 0.1309-0.3020i\\ 0.0626+0.0170i & 0.0132-0.0172i\end{pmatrix}.$$

The column vectors of $U-P_1$ are orthogonal to, and therefore are, scalar multiples of a single vector $x^2=(x_1^2, x_2^2)$:

$$x^2=\begin{pmatrix}0.0132-0.0172i\\ -0.0626-0.0170i\end{pmatrix}.$$

Similarly, the rows of $U-P_1$ are scalar multiples of

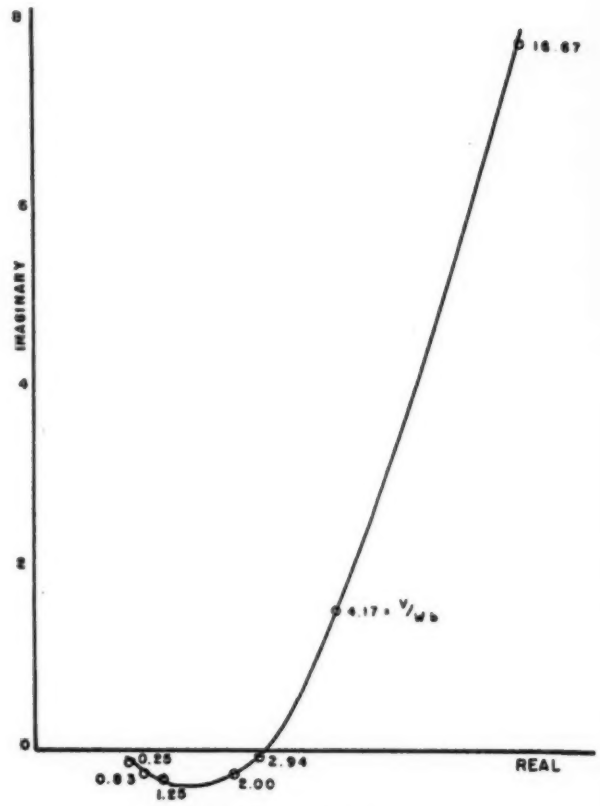


FIG. 2(b). Characteristic root λ_2.

$y^2=(y_1^2, y_2^2)$:

$$y^2=(0.0132-0.0172i,\ -0.1309+0.3020i);$$

and x^2 and y^2 are scalar multiples of e^2 and f^2, respectively:

$$\lambda_2=y^2Ax^2/(y^2, x^2)$$

$$=0.0262\angle -67°29'/0.0216\angle -52°20'$$

$$=1.21\angle -15°9'=1.17-0.32i.$$

As a numerical check, compare $\lambda_1+\lambda_2$ with $Tr(A)$:

$$\lambda_1+\lambda_2=20.77-4.22i,$$

$$Tr(A)=20.79-4.22i.$$

This calculation has been repeated for several values of $v/\omega b$ and the characteristic values λ plotted in Fig. 2.

To obtain the flutter characteristics of the three degree of freedom case, one calculates the terms on the right side of Eq. (5.5) for $n=2$:

$$(a_{n+1}, e_r^1)\cdot(f_s^1a_{s,n+1})=[a_{3,r}v_{xr}]\cdot[v_{ys}a_{s,3}]/(v_y, v_x)=$$

$$(0.02527\angle -17°49')(1.83391\angle -157°55')/1.013\angle -1°=0.0457\angle -174°43',$$

$$(a_{n+1,r}e_r^2)(f_s^2a_{s,n+1})=[a_{3,r}x_r^2]\cdot[y_s^2a_{s,3}]/(y^2, x^2)$$

$$=(0.00224\angle 134°49')(0.05668\angle -23°12')/0.216\angle -52°20'$$

$$=0.00602\angle 164°58',$$

$$f(\lambda)=-(0.04574\angle -174°43'/19.60-3.90i-\lambda)-(0.00602\angle 164°58'/1.17-0.32i-\lambda)$$

$$+(-0.007147-0.018058-0.00045)i/\lambda.$$

Figure 3 is a plot of $f(\lambda)$ for λ real. The aileron is seen to couple more strongly with the second mode in the sense that the response function is closest to zero for values of λ of the order of magnitude of $R(\lambda_2)$. The second mode is predominantly torsion so that one can say that in the three degree-of-freedom case, aileron deflection is strongly coupled with wing torsion.

The imaginary part of λ can be readily estimated by scaling the closest distance from the $f(\lambda)$ curve to establish the local scale factor. One obtains, for $v/\omega b=0.83$,

$$\text{Imag pt } \lambda=+0.002.$$

Comparison of the curves show that the response function has crossed the origin. One obtains in this way a crossing at $v/\omega b=0.82$, $\lambda=1.08$, which compares favorably with the corresponding values $v/\omega b=0.875$, $\lambda=1.06$ reported in reference 5.

APPENDIX. DERIVATION OF RESPONSE FUNCTION

The homogeneous Eqs. (2.7) and (2.12) may be written in the form of matrix equations:

$$(A-\lambda B)x=0, \tag{A1}$$

$$y(A-\lambda B)=0. \tag{A2}$$

Write the nonhomogeneous Eqs. (2.9) in the form

$$(A-\lambda B)x=w. \tag{A3}$$

Here x and w represent matrices of one column and y a matrix of one row.

Let λ_j represent a characteristic root, and let e^j and f^j represent corresponding characteristic vectors which are solutions of Eqs. (A1) and (A2), respectively. From the definition, one obtains

$$(A-\lambda_j B)e^j=0, \quad f^k(A-\lambda_k B)=0, \quad j,k=1,2,\cdots. \tag{A4}$$

Hence, it follows that

$$f^k(A-\lambda_j B)e^j=0, \quad f^k(A-\lambda_k B)e^j=0. \tag{A5}$$

Subtracting Eqs. (A5), one obtains

$$(\lambda_k-\lambda_j)f^k B e^j=0.$$

Therefore, if $\lambda_k\neq\lambda_j$, it follows that

$$f^k B e^j=0, \quad \lambda_k\neq\lambda_j. \tag{A6}$$

and from (A5),

$$f^k A e^j=0, \quad \lambda_k\neq\lambda_j. \tag{A7}$$

We shall assume for the moment that (a) there are n characteristic roots λ_j and that they are all distinct and (b) $f^j B e^j\neq 0$, $j=1,2,\cdots,n$. By suitable normalization, it follows that there are n direct characteristic modes e^j and n conjugate modes f^j such that

$$f^k A e^j=\lambda_j\delta_{kj}, \quad f^k B e^j=\delta_{kj}, \quad \text{(no summations)}, \tag{A8}$$

where δ_{kj} is the Kronecker delta.

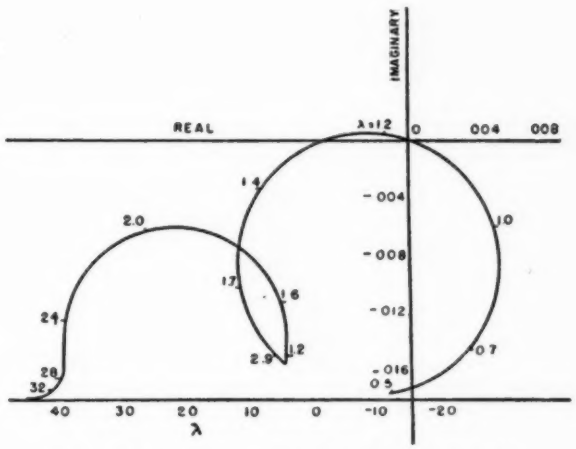


FIG. 3. Aileron response function P/β $(v/b\omega=0.83)$.

Now consider the solution of the nonhomogeneous Eq. (A3). Assume that every x may be written as a linear combination of the e^j,

$$x=\sum_{j=1}^{n}\xi_j e^j. \tag{A9}$$

Substitution into Eq. (A3) followed by left multiplication of f^k gives

$$f^k(A-\lambda B)\sum\xi_j e^j=f^k w.$$

From the orthogonality relations (A8), one obtains

$$(\lambda_k-\lambda)\xi_k=f^k w.$$

Solving for ξ_k and substituting into (A9) gives

$$x=\Sigma[(f^k w)/(\lambda_k-\lambda)]e^k. \tag{A10}$$

In the more general case in which Eq. (A9) does not hold, it can be shown with several additional assumptions that

$$x=\sum_{j=1}^{m}\xi_j e^j+g, \tag{A11}$$

where $m(<n)$ is the rank of matrix B and g is a mode which develops no potential energy (called a *free* mode). That is, in matrix terminology, we have

$$Bg=0. \tag{A12}$$

The solution of Eq. (A2) then takes the form

$$x=\sum_{k=1}^{m}\frac{(f^k w)}{\lambda_k-\lambda}e^k+g, \tag{A13}$$

where g is the free body response to the external force obtained by solution of

$$\Sigma' a_{rs}x_s=w_r, \tag{A14}$$

where Σ' indicates summation over the free modes of $x_1, x_2, \cdots, x_n$.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

The Susceptance of a Thin Iris in Circular Wave Guide with the TM_{01} Mode Incident

KENNETH L. DUNNING AND RUFUS G. FELLERS
Naval Research Laboratory, Washington, D. C.
(Received February 5, 1951)

Certain wave-guide boundary value problems can be formulated in terms of lumped-constant circuits and distributed-constant transmission lines. Equivalent circuit voltages and currents can be introduced as measures of the transverse electric and magnetic fields. Making use of these concepts and Schwinger's integral equation method, the susceptance of a thin circular iris in circular cylindrical wave guide with the TM_{01} mode incident is discussed and calculated. Results are compared with experimental data.

INTRODUCTION

THE impedance concept has been employed in a number of analyses treating obstacles and discontinuities in wave guides and parallel-plate transmission lines.[1–5] The results are usually cast in such a form that the behavior of dominant mode fields in regions remote from the obstacle or discontinuity can be predicted by considering a distributed-constant transmission line and a lumped-constant network. In the case of an obstacle placed in an otherwise uniform wave guide, the wave guide may be represented by a single transmission line and the obstacle may be represented by a four-terminal network. Marcuvitz has generalized this approach to the case of N interconnected wave guides and a $2N$ terminal network.[4] Schwinger's integral equation method enables one to evaluate the parameters of the obstacle network.

In practice, the guide cross section is usually arranged so that only one mode known as the dominant mode propagates, the remaining modes being damped exponentially or not excited. The introduction of an obstacle in a guide of otherwise uniform cross section gives rise to such modes as are necessary to satisfy boundary conditions on the obstacle. If all excited modes except the dominant one are exponentially damped, the detectable energy at points along the guide sufficiently remote from the point of excitation is entirely in the dominant mode. In practice, "sufficiently remote" implies a distance of the order of a wavelength.

The dominant-mode transverse electric field may be represented by a transmission-line voltage and the dominant-mode transverse magnetic field may be represented by a transmission-line current. Marcuvitz points out that, although the circuit relations here involved appear analogous to low frequency circuit relations, a distinction must be made.[6] In the type of problem being considered presently, the voltages and currents are measures of the "far" electric and magnetic fields; in low frequency problems, the voltages and currents are measures of the surface values of electric and magnetic fields in the immediate vicinity of the discontinuity structure. In both cases, however, the quantities represented are those most readily measurable.

The analysis at hand will be developed in the following order—using mks units. The mode functions which describe the electromagnetic field in a hollow uniform circular metal cylinder will be discussed and only the transverse fields will be utilized henceforth. Next, the modes generated by the introduction of a thin metal iris will be considered. The problem will then be reduced to one of a single dimension by defining voltages and currents to represent dominant-mode transverse electric and magnetic fields. After discussing the form that the network representing the iris must take, the problem will be formulated first in terms of boundary conditions in the aperture and second in terms of boundary conditions on the iris. Each of these formulations leads to a result. The first leads to an upper bound for the iris susceptance; the second leads to a lower bound for iris susceptance. Finally, the analytical results will be compared with results obtained experimentally.

In a hollow metal circular cylinder the TE_{11} mode has the lowest cut-off frequency. The TM_{01} mode has the next lowest cut-off frequency. The TE_{11} mode is usually considered to be the dominant mode; however, if the geometry of the exciting terminal of a circular wave guide has proper symmetry, the TE_{11} mode will not be excited and the TM_{01} may be termed dominant.

FIELDS IN A HOLLOW CIRCULAR METAL CYLINDER

Upon assuming the time variation of the electric and magnetic fields in a medium of constant permeability μ, dielectric constant ϵ and zero conductivity to be determined by a factor $e^{j\omega t}$, where $j=\sqrt{-1}$, it follows immediately from Maxwell's equations that the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ must satisfy the vector

[1] S. A. Schelkunoff, Quart. App. Math. **1**, 78 (1943).
[2] J. R. Whinnery and H. W. Jamison, Proc. Inst. Radio Engrs. **32**, 98 (1944).
[3] J. S. Schwinger, "Notes on lectures by Julian Schwinger" (unpublished).
[4] N. Marcuvitz, "Wave guide circuit theory: coupling of wave guides by small apertures," Polytechnic Institute of Brooklyn Report No. R 157–47, PIB-106.
[5] John W. Miles, Proc. Inst. Radio Engrs. **34**, 728 (1946).
[6] See reference 4, p. 3a.

wave equation

$$\nabla^2\mathbf{E}+k^2\mathbf{E}=0 \tag{1}$$

$$\nabla^2\mathbf{H}+k^2\mathbf{H}=0, \tag{2}$$

where $k^2=\omega^2\mu\epsilon$ and k is the phase constant of free space, ω being the angular frequency. If wave propagation in a z direction only is to be dealt with, the z dependence of the fields may be expressed in a factor $e^{\pm\gamma z}$ where γ is a complex propagation constant, $\gamma=\alpha+j\beta$, α being the attenuation constant (which vanishes for lossless media and boundaries) and β being the phase constant; α and β are positive numbers. The minus sign indicates transmission in a plus z direction; the plus sign indicates transmission in a minus z direction.

If the wave equations (1) and (2) are solved subject to the boundary condition that a conductor of infinite conductivity be at $r=a$ (see Fig. 1), an infinite orthogonal set of solutions or mode functions will be found. These solutions are well known.

Only the transverse component of these solutions is of interest[7] and is designated by $\mathbf{\Phi}e^{j\omega t-\gamma z}$ where

$$\mathbf{\Phi}=\mathbf{r}_1\phi_r+\boldsymbol{\theta}_1\phi_\theta, \tag{3}$$

$\mathbf{\Phi}$ is a solution of the vector wave equation.

$$\nabla^2\mathbf{\Phi}+(k^2+\gamma^2)\mathbf{\Phi}=0. \tag{4}$$

Solutions to the wave equation fall into two classes, one being termed transverse electric (TE), the other transverse magnetic (TM). Propagation in the TM modes is characterized by

$$k^2-\beta_{mn}{}^2=(\rho_{mn}/a)^2 \tag{5}$$

where ρ_{mn} is the nth root of $J_m(x)$. The total transverse field that can exist in a wave guide of uniform cross section can be expressed as a sum of mode functions

$$\mathbf{E}(r,\theta,z,t)=\sum_{mn}C_{mn}\mathbf{\Phi}_{mn}\exp[j(\omega t\pm\beta_{mn}z)], \tag{6}$$

where C_{mn} is an amplitude factor associated with the mnth mode.

If the radius of the wave guide is fixed at a value which permits only the TE_{11} and the TM_{01} modes to propagate and if, in addition, the method of excitation is such that the TE_{11} mode will not be set up, only the TM_{01} mode will exist in a guide of uniform cross section. This mode is axially symmetric having radial electric lines of force and axially concentric magnetic lines of force. It will be referred to as the dominant mode.

If an obstacle is introduced into a wave guide in which the dominant mode is propagating, other modes will be excited. The fields at any point (except perhaps in the section occupied by the obstacle) may be described by (6), the amplitudes of the new modes being determined by that of the incident dominant mode and the nature of the obstacle. The incident wave causes current to flow in the obstacle, and new electric and magnetic fields are created to satisfy boundary conditions.

[7] See reference 5, p. 729.

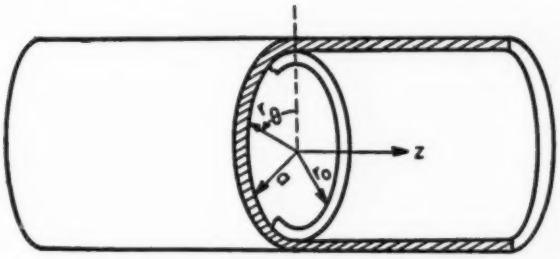


FIG. 1. Wave-guide coordinate system and iris.

In general, the modes excited by an obstacle will be of both the TM and TE types. However, if the obstacle is an infinitely thin, perfectly conducting, planar, axially symmetric iris with a circular aperture (see Fig. 1), only axially symmetric modes with no angular component of electric field will be excited. The excited modes will be axially symmetric because the TM_{01} electric exciting field and the geometry of the iris are independent of the angular variable θ. They will have no electric component in the $\boldsymbol{\theta}_1$ direction because there is none in the exciting field. In view of these considerations, an examination of the uniform-guide solutions which occur throughout the literature will show that only TM_{0n} waves can exist. The mode functions which are reproduced here for convenience will then have the following form:

$$\begin{aligned}
E_r&=j\beta_{0n}\frac{a}{\rho_{0n}}J_1\left(\frac{\rho_{0n}}{a}r\right)\exp[j(\omega t\pm\beta_{0n}z)],\\
E_\theta&=0,\\
E_z&=J_0\left(\frac{\rho_{0n}}{a}r\right)\exp[j(\omega t\pm\beta_{0n}z)],\\
H_r&=0,\\
H_\theta&=j\omega\epsilon\left(\frac{a}{\rho_{0n}}\right)J_1\left(\frac{\rho_{0n}}{a}r\right)\exp[j(\omega t\pm\beta_{0n}z)],\\
H_z&=0.
\end{aligned} \tag{7}$$

TRANSMISSION LINE CONCEPTS AND EQUIVALENT CIRCUIT

The transverse electric and magnetic fields for each mode can be represented by a voltage and a current respectively on an equivalent distributed-constant transmission line having the phase constant of the mode represented. Let a transmission line voltage $V_{mn}(z)$ and a transmission line current $I_{mn}(z)$ with positive directions as shown in Fig. 2 be defined in terms of the transverse electric field $\mathbf{E}_{mn}$ and a transverse magnetic field $\mathbf{H}_{mn}$ by the following equations:

$$\mathbf{E}_{mn}(r,\theta,z,t)=AV_{mn}(z,t)\mathbf{\Phi}_{mn}(r,\theta), \tag{8}$$

$$\mathbf{H}_{mn}(r,\theta,z,t)=BI_{mn}(z,t)\mathbf{z}_1\times\mathbf{\Phi}_{mn}(r,\theta), \tag{9}$$

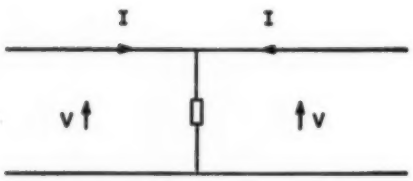


FIG. 2. Shunt element representing an iris.

where A and B are arbitrary constants. Note that the voltage V_{mn} is proportional to the amplitude of the transverse electric field of the TM_{mn} mode and that the current I_{mn} is proportional to the amplitude of the transverse magnetic field of this mode.

It is possible to take into account the effect of a thin axially symmetric iris on the dominant mode transverse fields by representing it on the equivalent line by a shunt susceptance. This equivalence is a consequence of the fact that the iris is planar, is of infinitesimal thickness, and is lossless.

Miles has derived expressions for the susceptance of an infinitely thin iris at the junction between two cylindrical guides of arbitrary cross sections and cast them in a form used by Schwinger.

The expressions for the value of this susceptance are obtained by substituting the correct quantities for the TM_{01} case in Mile's expressions. Use of his expression (34)[8] yields the result below, in which $\mathfrak{E}(r)$ represents the field in the aperture.

$$B_{\mathfrak{E}}=\frac{2\beta J_1^2(\rho_{01})\sum_{n=2}^{\infty}\frac{1}{J_1^2(\rho_{0n})[(\rho_{0n}/a)^2-k^2]^{\frac{1}{2}}}\left[\int_0^{r_0} rJ_1(\rho_{0n}r/a)\mathfrak{E}(r)dr\right]^2}{\left[\int_0^{r_0} rJ_1(\rho_{01}r/a)\mathfrak{E}(r)dr\right]^2}. \tag{10}$$

Since the right member of (10) is positive, $B_{\mathfrak{E}}$ is capacitive. Schwinger has shown that an expression of the form of (10) is stationary with respect to arbitrarily small variations of $\mathfrak{E}$ about its true value. Further, he has shown that if $\mathfrak{E}$ is assumed to be any function of r whatsoever, the value of $B_{\mathfrak{E}}$ resulting will always be greater than the true value.

Computations have been made assuming $\mathfrak{E}$ to be a two-term power series and the findings plotted against experimental data in Fig. 3. These results will be discussed later. Next, an integral equation associated with a lower bound for susceptance will be formed. Use of Miles Eq. (53) yields the expression for susceptance below in terms of $K(r)$, the current on the iris. This equation gives a lower bound to the true value of susceptance.

$$\frac{1}{B_K}=\frac{\frac{J_1^2(\rho_{01})}{2\beta}\sum_{n=2}^{\infty}\frac{[(\rho_{0n}/a)^2-k^2]^{\frac{1}{2}}}{J_1^2(\rho_{0n})}\left[\int_{r_0}^{a} rK(r)J_1(\rho_{0n}r/a)dr\right]^2}{\left[\int_{r_0}^{a} rK(r)J_1(\rho_{01}r/a)dr\right]^2}. \tag{11}$$

CALCULATION OF SUSCEPTANCE-ASSUMING ELECTRIC FIELD IN THE APERTURE

To obtain an expression from Eq. (10) from which numerical values of B may be calculated, a function $\mathfrak{E}(r)$ may be assumed. A function $\mathfrak{E}(r)$ has been selected which will yield accurate values for B for large apertures—without excessively tedious calculation.

Inspection of TM_{0n} mode functions reveals that all transverse electric fields vanish at $r=0$. A field given by

$$\mathfrak{E}(r)=\sum_{m=1}^{\infty} A_m r^{2m-1} \quad m=1, 2, 3, \cdots \tag{12}$$

will satisfy this requirement and leads to expressions easily integrated. Equation (10) then becomes

$$B_{\mathfrak{E}}=\frac{2\beta J_1^2(\rho_{01})\sum_{n=2}^{\infty}\frac{1}{J_1^2(\rho_{0n})[(\rho_{0n}/a)^2-k^2]^{\frac{1}{2}}}\left[\int_0^{r_0} rJ_1(\rho_{0n}r/a)\sum_{m=1}^{\infty} A_m r^{2m-1}dr\right]^2}{\left[\int_0^{r_0} rJ_1(\rho_{01}r/a)\sum_{m=1}^{\infty} A_m r^{2m-1}dr\right]^2}. \tag{13}$$

[8] See reference 5, pp. 732, 734.

To simplify notation, quantities may be defined as follows

$$D_{mp}=D_{pm}=2\beta J_1^2(\rho_{01})\sum_{n=2}^{\infty}\frac{1}{J_1^2(\rho_{0n})[(\rho_{0n}/a)^2-k^2]^{\frac{1}{2}}}\left[\int_0^{r_0} rJ_1(\rho_{0n}r/a)r^{2p-1}dr\right]\left[\int_0^{r_0} rJ_1(\rho_{0n}r/a)r^{2m-1}dr\right], \quad (14)$$

$$C_m=\int_0^{r_0} rJ_1(\rho_{0n}r/a)r^{2m-1}dr. \quad (15)$$

Equation (13) then becomes

$$B_{\mathfrak{E}}=\frac{\sum_{p,m=1}^{\infty} D_{pm}A_pA_m}{\left[\sum_{m=1}^{\infty} C_mA_m\right]^2}. \quad (13a)$$

This may be written

$$B_{\mathfrak{E}}\sum_{p,m=1}^{\infty} C_pC_mA_pA_m=\sum_{p,m=1}^{\infty} D_{pm}A_pA_m. \quad (13b)$$

The requirement that $B_{\mathfrak{E}}$ be stationary with respect to arbitrarily small variations of $\mathfrak{E}(r)$ means that the derivative of $B_{\mathfrak{E}}$ with respect to any of the coefficients A_m must be zero. Differentiating and using this fact,

$$B_{\mathfrak{E}}\sum_{p=1}^{\infty} C_pC_mA_p=\sum_{p=1}^{\infty} D_{pm}A_p. \quad (16)$$

Transposing and collecting,

$$\sum_{p=1}^{\infty} A_p(D_{pm}-B_{\mathfrak{E}}C_mC_p)=0. \quad (17)$$

This is an infinite set of linear homogeneous equations in the unknown expansion coefficients A_p. In order for a nonzero solution to exist, the determinant of coefficients must vanish.

Since this resulting equation is linear in $B_{\mathfrak{E}}$, $B_{\mathfrak{E}}$ has one and only one value. Because of the stationary character of $B_{\mathfrak{E}}$ with respect to $\mathfrak{E}(r)$, the error in $B_{\mathfrak{E}}$ produced by an incorrect field function will be proportional to the square of the difference between $\mathfrak{E}(r)$ and the correct field function. This speeds calculation if the values assumed initially for $\mathfrak{E}(r)$ are much in error.

If the determinant is solved for $B_{\mathfrak{E}}$—assuming that all coefficients of (12) except the first are zero,

$$B_{\mathfrak{E}}=D_{11}/C_1^2. \quad (18)$$

If the determinant is solved for $B_{\mathfrak{E}}$—assuming that all coefficients of (12) except the first two are zero,

$$B_{\mathfrak{E}}=\frac{D_{11}/C_1^2\cdot D_{22}/C_2^2-(D_{12}/C_1C_2)^2}{D_{11}/C_1^2+D_{22}/C_2^2-2D_{12}/C_1C_2}. \quad (19)$$

The results of calculations for a particular wave guide and wavelength using formulas (18) and (19) are plotted in Fig. 3. The wave guide radius is 1.99 cm; the air wavelength is 4.5 cm. The agreement between calculated values and experimental values is best for large apertures. For very large apertures formula (18) is of useful accuracy. The details for a calculation of one point are tabulated in Appendix I. The experimental determinations will be discussed in a later section.

CALCULATION OF SUSCEPTANCE ASSUMING A FUNCTION PROPORTIONAL TO CURRENT ON THE IRIS

By similar methods the equations below for the lower bound of susceptance are found. Equation (20) is obtained from a one-term current assumption and (21)

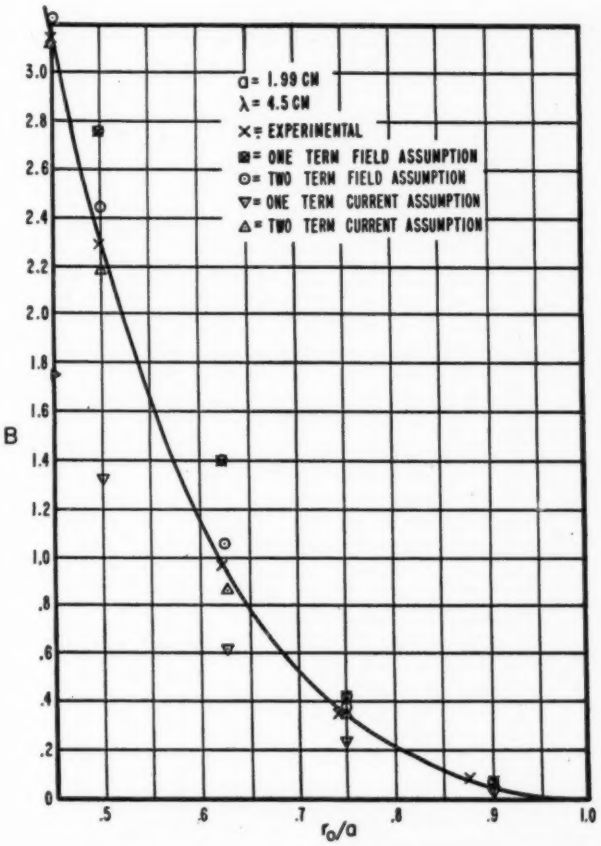


FIG. 3. Relation between susceptance and aperture-to-guide radius ratio.

TABLE I. Voltage standing wave ratio measurements and corresponding values of iris susceptance for several values of r_0/a.

r_0/a	*VSWR* with load adding	Reflection coefficient (R) corrected for load	$B=\frac{2\lvert R\rvert}{[1-\lvert R\rvert^2]^{\frac{1}{2}}}$
0.877	1.10	0.038	0.076
0.745	1.45	0.174	0.354
0.622	2.60	0.445	0.965
0.498	7.50	0.755	2.30
0.445	12.8	0.845	3.15

from a two-term current assumption.

$$1/B_K=\mathfrak{D}_{11}/\mathfrak{C}_1^2 \tag{20}$$

$$\frac{1}{B_K}=\frac{\mathfrak{D}_{11}/\mathfrak{C}_1^2\cdot\mathfrak{D}_{22}/\mathfrak{C}_2^2-(\mathfrak{D}_{12}/\mathfrak{C}_1\mathfrak{C}_2)^2}{\mathfrak{D}_{11}/\mathfrak{C}_1^2+\mathfrak{D}_{22}/\mathfrak{C}_2^2-2\mathfrak{D}_{12}/\mathfrak{C}_1\mathfrak{C}_2}. \tag{21}$$

MEASUREMENT OF SUSCEPTANCE

To obtain experimental verification of calculated values, measurements were made on six irises at a wavelength of 4.5 cm in a brass pipe of 1.99 cm radius.

To insure that the energy incident upon the irises was almost entirely in the TM_{01} mode, a coaxial-to-circular wave-guide transition was constructed with a high degree of symmetry. Measurements revealed that the power in the TE_{11} mode was about 0.01 percent of that in the TM_{01} mode.

Standing wave measurements were made for irises of aperture-to-guide ratios as shown in Table I along with calculated values of normalized susceptance. The voltage standing wave ratio of the load was 1.02; this was taken into account in computing the susceptances.

TABLE II. Evaluation of quantities D_{11} etc. for various values of r_0/a.

r_0/a	D_{11}	D_{22}	D_{12}	C_1	C_2
0.444	0.0290	0.0063	0.0129	0.0841	0.0433
0.50	0.0474	0.0374	0.0270	0.131	0.0851
0.625	0.1013	0.0956	0.0920	0.2981	0.2996
0.75	0.1322	0.2442	0.1585	0.5626	0.8012
0.90	0.0593	0.5074	0.1019	1.0297	2.074

ACKNOWLEDGMENT

This work has been used as the basis for a thesis: Master of Science, University of Maryland (K. L. Dunning). Acknowledgment is here made to Dr. M. H. Johnson and Dr. John P. Hagen, Naval Research Laboratory for their helpful criticism in the preparation of this paper.

TABLE III. Evaluation of $B_{\mathfrak{C}}$ for several values of r_0/a.

r_0/a	$B_{\mathfrak{C}}$ (one-term field)	$B_{\mathfrak{C}}$ (two-term field)
0.445	4.10	3.2
0.50	2.76	2.44
0.625	1.40	1.056
0.75	0.417	0.371
0.90	0.0560	0.0551

APPENDIX I

Numerical Evaluation of $B_{\mathfrak{C}}$

In order to evaluate formulas (18) and (19), D_{11}, D_{12}, D_{22}, C_1, and C_2 are determined. To illustrate the manner in which the magnitude of these quantities change with r_0/a for a specific wave guide and wavelength, let $a=1.99$ cm and wavelength $=4.5$ cm. The values of these quantities for various aperture-to-guide ratios are tabulated in Table II.

Table III lists values of susceptance calculated from (18) and (19) using a one-term field assumption and two-term field assumption, respectively. These values are plotted in Fig. 3.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

Predominantly Subharmonic Oscillations

CARL A. LUDEKE
University of Cincinnati, Cincinnati, Ohio
(Received February 5, 1951)

In this paper an electromechanical analog previously reported, is used to investigate the phenomena of subharmonic oscillations. Special consideration is given to the case in which the subharmonic "predominates," i.e., in which the amplitude of the fourier component having the subharmonic frequency is greater than the amplitude of the fundamental. An investigation is then made of the possible transitions between subharmonics of different orders.

INTRODUCTION

ONE of the most interesting phenomena of nonlinear forced oscillations is the phenomena of subharmonic resonance or frequency demultiplication. In this phenomena a forcing function, of frequency ω, maintains oscillations whose frequencies are ω/n, where n is an integer greater than one. If these oscillations have amplitudes which are small compared to the amplitude of the oscillation whose frequency is ω, the subject has been discussed in the literature.[1,2] It seemed to this author, however, that the really interesting case occurred when the subharmonic oscillation predominated.[3,4]

THEORY

Consider, therefore, the differential equation

$$d^2x/dt^2+k_1x+k_3x^3=P_0\sin\omega t \tag{1}$$

and assume it has a solution

$$x=A\sin(\omega t/3)+pA\sin\omega t \tag{2}$$

in which p is so small that terms in p^2 can be neglected. Under what condition is this assumption possible? Upon substituting Eq. (2) in Eq. (1) and dropping terms of the order of p^2, we find

$$[(k_1-m\omega^2/9)A+3k_3A^3/4]\sin(\omega t/3)+\{(k_1-m\omega^2)pA+k_3A^3[-1/4+3p/2-(3p/2)\cos(2\omega t/3)]-P_0\}\sin\omega t=0. \tag{3}$$

Thus if $A\neq0$ and Eq. (2) is to be a solution of Eq. (1), we have

$$k_1-m\omega^2/9+3k_3A^2/4=0 \tag{4}$$

and

$$p=P_0/A(k_1-m\omega^2)-X(k_1-m\omega^2/9)/(k_1-m\omega^2) \tag{5}$$

where

$$X=(4/3)[-1/4+3p/2-(3p/2)\cos(2\omega t/3)]. \tag{6}$$

Since X is bounded in value, the last part of Eq. (5) can be made negligible by choosing large values of ω. For example, letting $\omega=q(k_1/m)^{\frac{1}{2}}$, Eq. (5) becomes

$$p=P_0/A(k_1-m\omega^2)-X(1-q^2/9)/(1-q^2). \tag{7}$$

For $q\geqslant5$, the term in X becomes of the order of p^2 and can be neglected so that Eq. (7) can be written

$$p=P_0/A(k_1-m\omega^2). \tag{8}$$

From Eq. (8) we see that p can be kept small by making P_0 small. Solving Eq. (8) for pA, the coefficient of the $\sin\omega t$ term in the solution, we have

$$pA=P_0(k_1-m\omega^2). \tag{9}$$

Thus if P_0 is small and $\omega\geqslant5(k/m)^{\frac{1}{2}}$, Eq. (2) can be considered an approximate solution to Eq. (1). The coefficients A and pA in Eq. (2) are determined by Eqs. (4) and (9) respectively.

An examination of Eqs. (4) and (9) shows that the amplitude of the subharmonic follows the frequency amplitude relationship of the free vibration starting in this case at $\omega=3(k_1/m)^{\frac{1}{2}}$ rather than at $\omega=(k_1/m)^{\frac{1}{2}}$; while the amplitude of the oscillation having the frequency of the forcing function follows the resonance curve of a linear system. By generalizing Eq. (1) so that it is in the form

$$d^2x/dt^2+k_1x+k_2x^2+\cdots k_nx^n=P_0\sin\omega t \tag{10}$$

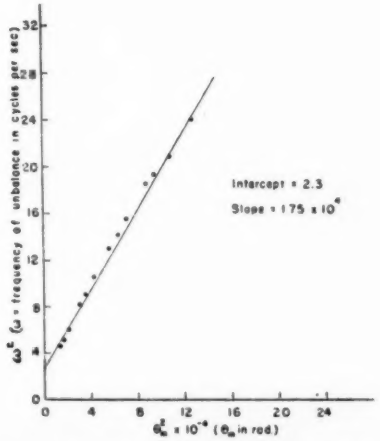


FIG. 1. ω^2 *versus* θ_m^2 for the free oscillation using springs NL 22 and L 1.

[1] K. O. Friedricks and J. J. Stoker, *Introduction to Non-Linear Mechanics* (8 lectures) incorporated in the set of twenty lectures organized under the Program of Advanced Instruction and Research in Mechanics at Brown University and delivered in the academic year 1942–1943.
[2] N. Minorsky, *Introduction to Non-Linear Mechanics* (J. W. Edwards, Ann Arbor, Michigan, 1947).
[3] C. A. Ludeke, J. Appl. Phys. **13**, 418–423 (1942).
[4] C. A. Ludeke, Am. J. Phys. **16**, 430–434 (1948).

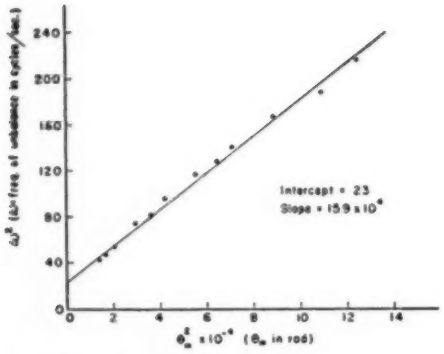


FIG. 2. ω^2 *versus* θ_m^2 for the forced subharmonic oscillation of order one-third using springs NL 22 and L 1.

we could expect solutions of the form

$$x=\sum_{2}^{n}A_{1/n}\sin(\omega t/n)+A_1\sin\omega t \tag{11}$$

in which, for those values of $A_{1/n}>A_1$ we know that the values of $A_{1/n}$ follow the frequency-amplitude relationship of the free vibration beginning at $\omega=n(k_1/m)^{\frac{1}{2}}$, and

$$A_1=P_0(k_1-m\omega^2) \tag{12}$$

where P_0 is small.

EXPERIMENTAL

By an apparatus previously described,[5] it was relatively easy to generate subharmonic oscillations, record the wave form, amplitude, and frequency, and check the preceding theory. The first experiments were performed on a system represented by the differential equation

$$Id^2\theta/dt^2+k_1\theta+k_3\theta^3=K_0\omega^2\sin\omega t \tag{13}$$

in which

$I=1.47\times10^4$ metric slug-cm^2
$\theta=$displacement in rad.
$t=$time in sec
$k_1=0.13\times10^7$ g cm/rad.
$k_3=11.08\times10^9$ g cm/rad.
$\omega=$angular frequency in rad./sec
$K\omega^2=$magnitude of the forcing function.

For free vibrations of a system represented by Eq. (13), we can use well-known methods to find that the frequency-amplitude relationship is

$$\omega^2=k_1/I+3k_3\theta_m^2/4I=2.24+1.43\times10^4\theta_m^2 \tag{14}$$

in which θ_m is the amplitude and ω is in cycles/sec. According to our theory we should find, therefore, that Eq. (13) has the solution

$$\theta=A_{\frac{1}{3}}\sin(\omega t/3)+A_1\sin\omega t \tag{15}$$

in which $A_{\frac{1}{3}}$ is to be found from the equation

$$(\omega/3)^2=2.24+1.43\times10^4A_{\frac{1}{3}}^2 \tag{16}$$

and A_1 is given by the equation

$$A_1=K\omega^2/(k_1-I\omega^2). \tag{17}$$

Since from a previous paper[6] we know that the coefficient of the term $k_3\theta_m^2/I$ in Eq. (14) is somewhat larger than $\frac{3}{4}$, it was decided to obtain first a frequency amplitude curve for the harmonic oscillation. This is shown in Fig. 1 and gives us the equation

$$\omega^2=2.3+1.75\times10^4\theta_m^2. \tag{18}$$

This corresponds to Eq. (14) with the coefficient of $k_3\theta_m^2/I$ increased from $\frac{3}{4}$ to 0.92. Using this experimentally determined value of the coefficient, Eq. (16) becomes

$$(\omega/3)^2=2.24+1.75\times10^4A_{\frac{1}{3}}^2. \tag{19}$$

As a first check on Eq. (19) a record was taken of a

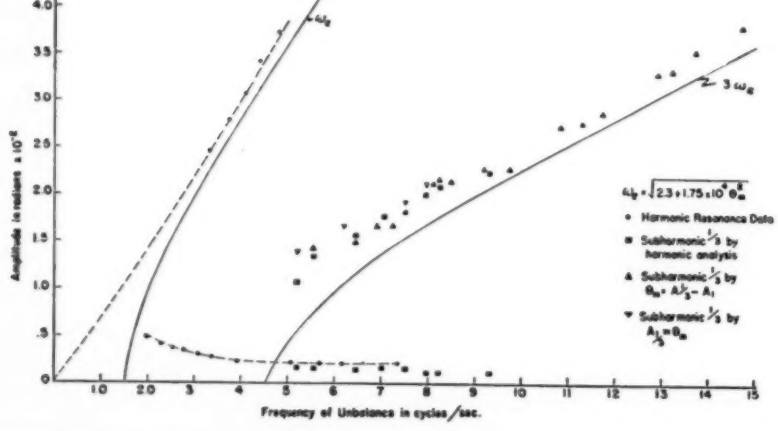


FIG. 3. Amplitude *versus* frequency for the harmonic oscillation and subharmonic oscillation of order one-third using springs NL 22 and L 1.

[5] C. A. Ludeke, J. Appl. Phys. **20**, 600–607 (1949).
[6] C. A. Ludeke, J. Appl. Phys. **20**, 694–699 (1949).

one-third subharmonic and the data was recorded assuming that A_1 of Eq. (15) was zero. This, of course, is not true, but it obviates the necessity of making harmonic analyses and gives a first approximation. The results are shown in Fig. 2 and lead to the equation

$$\omega^2=23+15.9\times10^4A_{\frac{1}{3}}^2 \tag{20}$$

which when divided by 9, becomes

$$(\omega/3)^2=2.5+1.77\times10^4A_{\frac{1}{3}}^2. \tag{21}$$

We note that Eq. (21) is in good agreement with Eq. (19).

With these results as encouragement other records were taken and harmonic analyses were made. Also use was made of the fact that if $A_{\frac{1}{3}}>3A$, Eq. (15) has a maximum given by

$$\theta_m=A_{\frac{1}{3}}-A_1, \tag{22}$$

and the value of A_1 is known from the harmonic resonance values beyond the jump frequency. The results are shown in Fig. 3.

The next set of experiments were performed with a spring system (S-1) whose torque deflection curve was

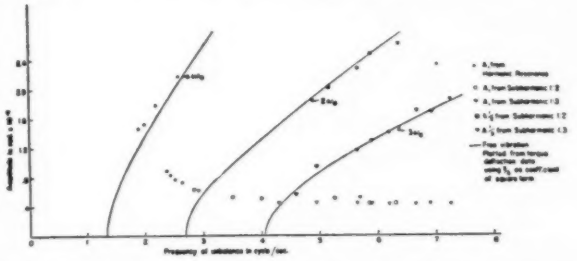


FIG. 4. Amplitude *versus* frequency for the harmonic oscillation and subharmonics of orders $\frac{1}{2}$ and $\frac{1}{3}$ using springs S-1.

of the form

$$T=k_0\theta+k_1\theta+k_2\theta^2+k_3\theta^3+k_4\theta^4+k_5\theta^5 \tag{23}$$

in which the torque T is measured in g cm and the deflection θ is in radians. The coefficients have the values

$$k_0=0.994\times10^4$$
$$k_1=0.106\times10^7$$
$$k_2=0.660\times10^8$$
$$k_3=10.9\times10^9$$
$$k_4=-5.5\times10^{10}$$
$$k_5=-55\times10^{11}.$$

With this torque-deflection curve the frequency amplitude relationship for the free vibration becomes

$$\omega^2=1.82+1.38\times10^4\theta_m^2-0.593\times10^7\theta_m^4 \tag{24}$$

in which θ_m is the amplitude in radians and ω is the frequency in cycles/sec.

With this spring system, subharmonics of orders $\frac{1}{2}$ and $\frac{1}{3}$ were recorded, and the resulting wave forms were harmonically analyzed. The results are shown in Fig. 4. In performing these experiments it became apparent

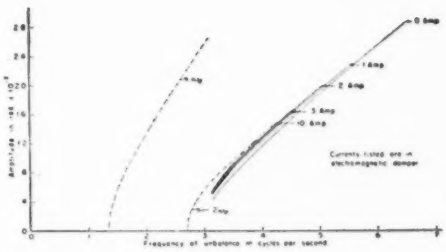


FIG. 5. The effect of viscous damping on the subharmonic of order one-half using springs S-1.

that any damping greatly reduced the chances of generating subharmonic oscillations; and also, that the effect of damping was most pronounced on the subharmonics of higher order. It was decided, therefore, to investigate the effect of damping on the lowest order subharmonic namely the one-half order. The data was taken by initiating the subharmonic at low frequencies and then increasing the frequency until the jump phenomena occurred. This procedure was repeated with various amounts of viscous damping introduced into the system by means of the electromagnetic damper. The results are shown in Fig. 5.

For further experiments a new spring system was utilized; it had a torque deflection curve of the form:

$$T=k_0+k_1\theta+k_2\theta^2+k_3\theta^3+k_4\theta^4+k_5\theta^5+k_6\theta^6+k_7\theta^7 \tag{25}$$

in which

$$k_0=0.60\times10^4$$
$$k_1=-0.75\times10^7$$
$$k_2=1.00\times10^8$$
$$k_3=4.15\times10^{10}$$
$$k_4=-1.35\times10^{11}$$
$$k_5=-3.36\times10^{13}$$
$$k_6=7.20\times10^{13}$$
$$k_7=1.29\times10^{16}.$$

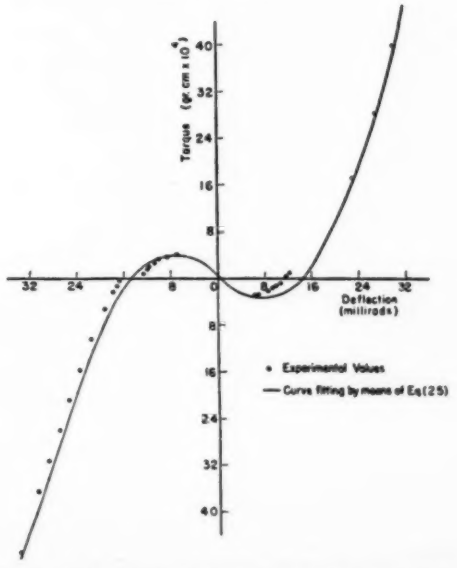


FIG. 6. The torque-deflection curve for springs S-4.

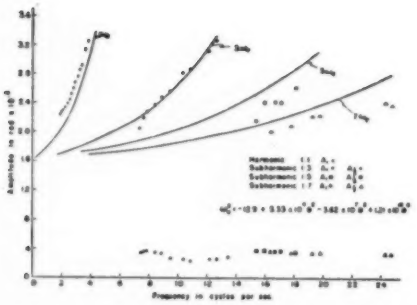


FIG. 7. Amplitude *versus* frequency for the harmonic oscillation and subharmonics of order $\frac{1}{3}$, $\frac{1}{5}$, and 1/7, using springs S-4.

This spring system is referred to as system S-4 and the torque deflection curve is shown in Fig. 6. Such a system has a frequency amplitude relationship for the free vibration of the form

$$\omega^2 = -12.9 + 5.33\times10^4\theta_m{}^2 - 3.62\times10^7\theta_m{}^4 + 1.21\times10^{10}\theta_m{}^6 \quad (26)$$

in which ω is in cycles/sec and θ_m is in radians. With this system it was possible to excite subharmonics of orders $\frac{1}{3}$, $\frac{1}{5}$, and 1/7th. The harmonic analyses of the resulting wave forms is shown in Fig. 7. All components with amplitudes less than 0.2×10^{-2} rad. were considered negligible.

One of the most interesting features of examining subharmonic solutions by means of an electromechanical analog, is the procedure necessary to obtain the subharmonics in the first place. If a nonlinear system such as those considered in this paper is excited by means of a rotating unbalance whose frequency is varied slowly from zero to a value beyond resonance frequency, the ordinary harmonic solution with the jump phenomena will be obtained. This author found, however, that if the frequency was increased rapidly through the "jump," the change in phase which takes place at the jump often caused the appearance of subharmonics. Such a case is shown in Fig. 8. After a subharmonic has been obtained, it is quite stable and the frequency can be varied slowly through a considerable range. In fact, the subharmonics themselves have jump phenomena and it is, therefore, possible to pass through these "jumps" in such a manner that beyond the "jump" a subharmonic of a different order is obtained. This is shown in Fig. 9.

In Fig. 8 the lower line of dots indicates when the rotating unbalances are horizontally away from the universal joints; in this position the forcing function is zero. The maximum amplitude for the case of harmonic resonance occurs at point A, in the interval A to B the phase shift occurs, and after point B we note that a

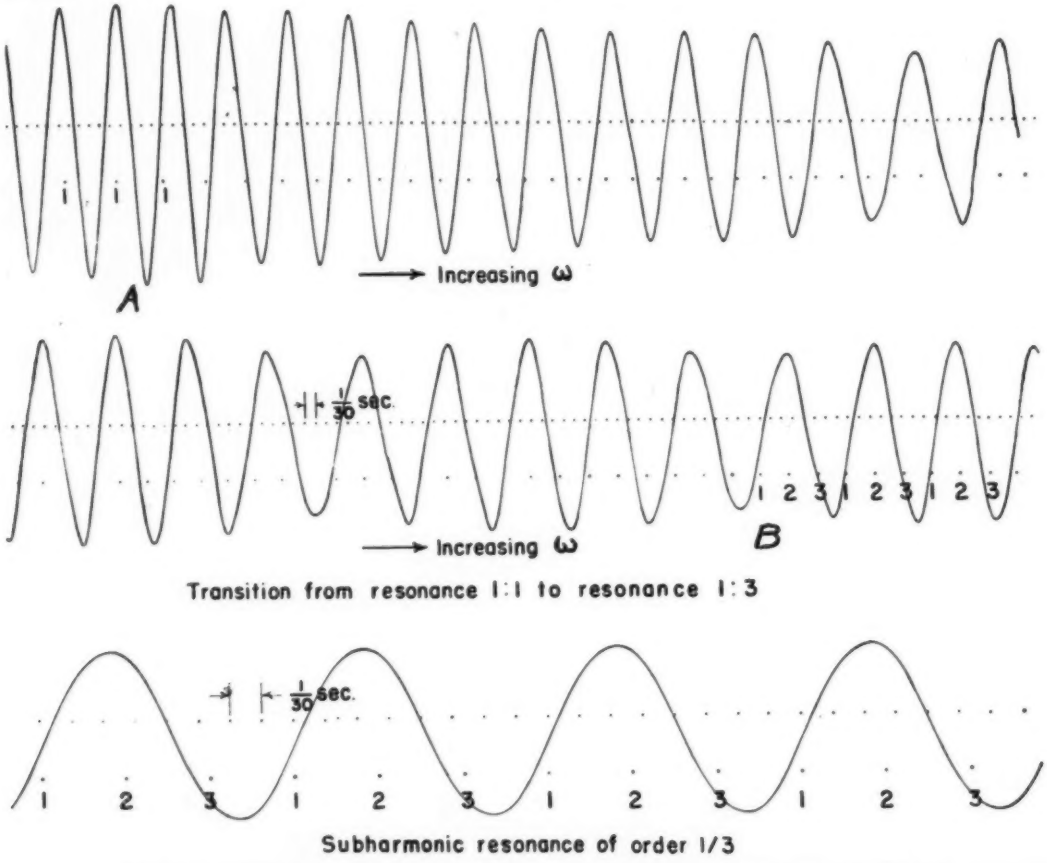

Increasing ω
A
1/30 sec.
1 2 3 1 2 3 1 2 3
Increasing ω
B
Transition from resonance 1:1 to resonance 1:3
1/30 sec.
1 2 3 1 2 3 1 2 3 1 2 3
Subharmonic resonance of order 1/3

FIG. 8. The transition from a harmonic oscillation to a subharmonic oscillation of order one-third generated by increasing ω.

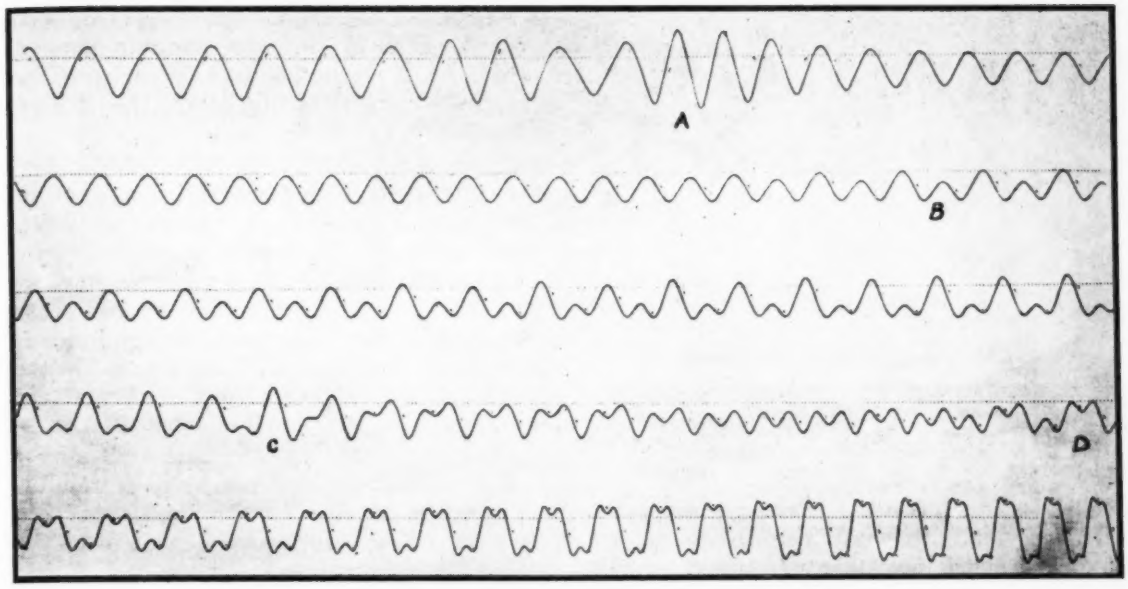


FIG. 9. Transition from harmonic to subharmonic of order one-half and then to subharmonic of order one-third.

subharmonic of order $\frac{1}{3}$rd has been obtained. The bottom part of Fig. 8 shows this same subharmonic considerably after the transition. The time scale has been increased in order that the wave form might be more easily examined.

In Fig. 9 the light used to record the phase dot was reflected by a mirror oscillating with the analog beam. This accounts for the vertical displacement of the phase spot. It still appears, however, when the rotating unbalances are horizontally away from the universal joint. Point A indicates the maximum amplitude for harmonic oscillations, the jump then occurs with its

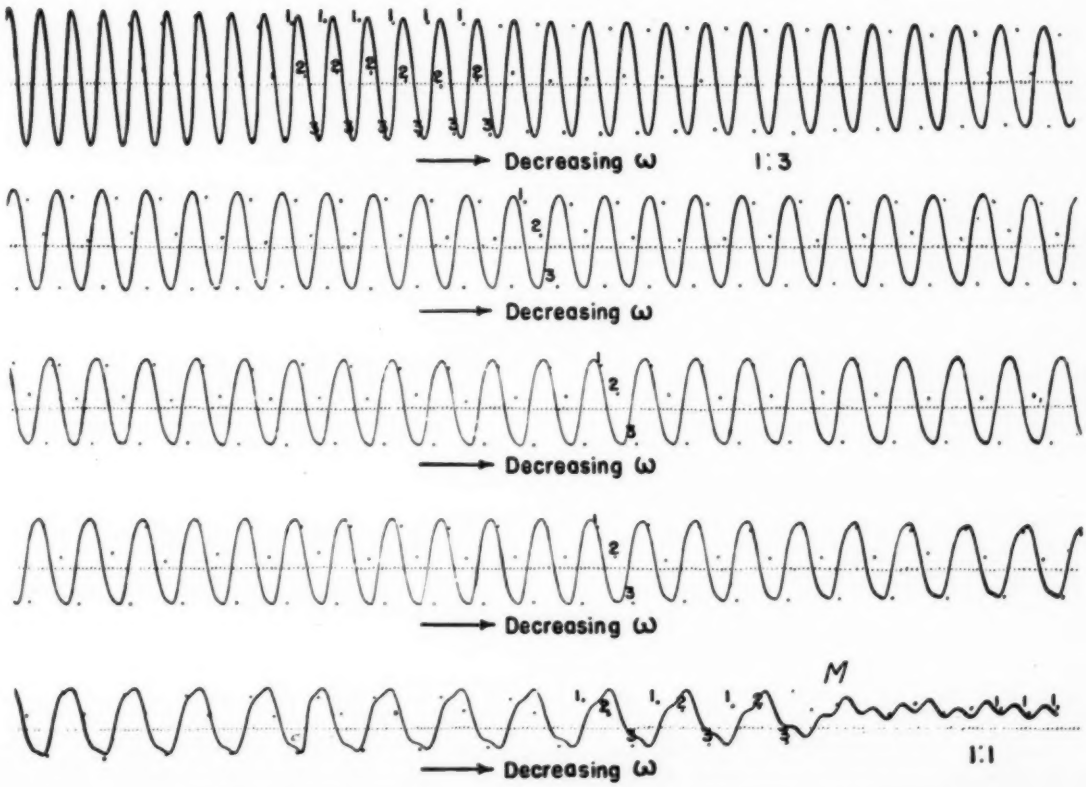


FIG. 10. The effect of decreasing ω on the wave form of a subharmonic of order one-third.

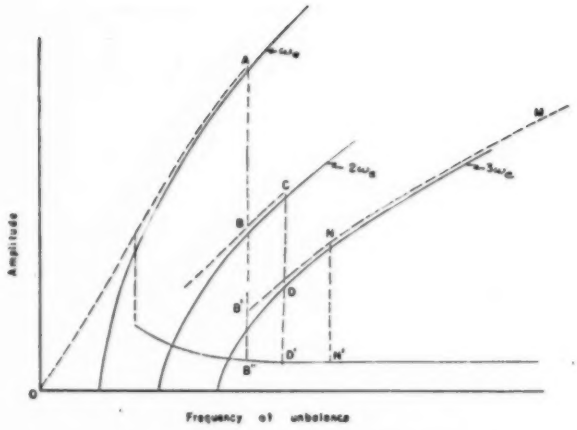


FIG. 11. A diagram of subharmonic transitions.

corresponding shift in phase and decrease in amplitude. As the frequency is continuously increased the motion changes so that point B indicates a subharmonic oscillation of order one-half. Further increase in frequency increases the amplitude of this subharmonic until at the point C the jump phenomena again takes place with a corresponding shift in phase and decrease in amplitude. A transition period occurs from C to D, but by time point D has been reached it is apparent that a subharmonic of order $\frac{1}{3}$rd has been obtained. By increasing the frequency the amplitude of this subharmonic is increased but the jump is never reached because the amplitude becomes too large to record.

In Fig. 10 we have an interesting record of the effect of gradually decreasing the frequency of a subharmonic oscillation of order $\frac{1}{3}$rd. The change in amplitude and wave form is clearly shown together with the jump (point M) to a harmonic oscillation of very small amplitude.

Figures 8, 9, and 10 give a fairly clear picture of just how transitions between subharmonics of different orders occur. Consider for example the resonance curves drawn in Fig. 11. Suppose as ω is increased we move along the harmonic resonance curve OA until we reach the peak at "A" and the jump occurs; the case of ordinary resonance requires that the amplitude decrease immediately to the point B'' and then follow the curve $B''D'N'$. However, because of the unstable condition during the jump, the amplitude may stop at B or B', resulting in subharmonic oscillations of order $\frac{1}{2}$ or $\frac{1}{3}$. In Fig. 8 the jump was from A to B' resulting in a subharmonic of order $\frac{1}{3}$rd. In Fig. 9 the first jump was from A to B ($\frac{1}{2}$ order), whereupon further increase in ω made the amplitude follow the curve B and C, at which point a second jump occurred (C to D) resulting in a subharmonic of order $\frac{1}{3}$rd. The amplitude then followed the curve DNM. In Fig. 10 the frequency of the subharmonic decreased and the amplitude followed the curve M to N at which point the jump was from N to N' resulting in a harmonic oscillation.

ACKNOWLEDGMENTS

It is a pleasure to thank the Research Corporation for the Frederick Gardner Cottrell Grant-in-Aid which made this research possible.

Erratum: On the Theory of Random Noise. Phenomenological Models. I

[J. Appl. Phys. 22, 1143 (1951)]

DAVID MIDDLETON
Cruft Laboratory, Harvard University, Cambridge, Massachusetts

EQUATIONS (2.32), (2.34), and (2.35) are incorrect. On the basis of (2.28) one must include not only the contributions of the overlapping impulses, but the two cases when there is no overlapping of the individual pulses in a typical pair. Accordingly, the correct form of (2.32) is

$$F_2(\xi_1,\xi_2;t)=\exp\left\{\gamma_{III}\int_{-\infty}^{\infty}w_1(a)da\int_{|t|}^{\infty}w_2(r)dr\int_0^{\beta(r-|t|)}\times[\exp ia(\xi_1+\xi_2)-1]dx+\int_{-\infty}^{\infty}w_1(a)da\int_0^{\infty}w_2(r)dr\times\left[\int_0^{\beta r}dx\{\exp ia\xi_1-1+\exp ia\xi_t-1\}\right]\right\} \quad (2.32a)$$

$$=F_1(\xi_1)F_1(\xi_2)\exp\left[-\gamma_{III}\rho(t)+\rho(t)\gamma_{III}\times\int_{-\infty}^{\infty}w_1(a)\exp ia(\xi_1+\xi_2)da\right], \quad (2.32b)$$

where $\rho(t)$ is correctly given by (2.33). Note that now as $t\to\infty$, $\rho(t)\to 0$ and $F_2(\xi_1,\xi_2;\infty)=F_1(\xi_1)F_1(\xi_2)$ as expected; in the limit one has a purely random process. For the statistics of a assumed in the previous paper, the characteristic function (2.32) becomes explicitly

$$F_2(\xi_1,\xi_2;t)=F_1(\xi_1)F_1(\xi_2)\exp\left[-\gamma_{III}\rho(t)\right]\sum_{n=0}^{\infty}\frac{[\gamma_{III}\rho(t)]^n}{n!}\exp\left[i\langle a\rangle n(\xi_1+\xi_2)-n\sigma^2(\xi_1+\xi_2)^2/2\right], \quad (2.34)$$

where $F_1(\xi)$ is given by (2.30). By inversion one finds finally for the corresponding probability density

$$W_2(X_1,X_2;t)=\exp\{-\gamma_{III}[2+\rho(t)]\}\Big[\delta(X_1-0)\delta(X_2-0)+\sum_{l+m+n>0}^{\infty}\exp\{-[(m+n)\langle a\rangle-X_2]^2/2\sigma^2(m+n)\}\cdot\exp\left\{-[l\langle a\rangle-X_1+nX_2/(m+n)]^2\Big/\left(\frac{mn+nl+ml}{m+n}\right)2\sigma^2\right\}\cdot\frac{[\gamma_{III}\rho(t)]^n}{l!m!n!}(2\pi\sigma^2)^{-1}(ml+mn+ln)^{-\frac{1}{2}}\Big]. \quad (2.35)$$

Some other corrections:

Equation (1.3), $W_2(X_2, Y_2, Z_2, \cdots)$ should read $W_1(X_2, Y_2, Z_2\cdots)$.

Reference 3, Paper II, 22, 1153 (1951) should read "James, Nichols, Phillips. . . ."

The Influence of Magnetic Fields upon the Propagation of Electromagnetic Waves in Artificial Dielectrics

E. R. WICHER*
Department of Physics, University of Wisconsin,† Madison, Wisconsin
(Received March 14, 1951)

An effect corresponding to the Faraday effect in natural dielectrics is predicted for a class of artificial dielectrics because of the existence of a Hall effect in the metallic components of the structure. A formula for Verdet's constant as a function of element polarizability and Hall coefficient is obtained. The resonance shift to be expected in a cavity resonator, filled with an artificial dielectric, and subjected to a strong magnetic field, is calculated. Prospects for the application of these effects are briefly considered.

I. GENERAL CONSIDERATIONS

SUBJECT to certain conditions, the dielectric constant of an artificial dielectric[1] may be expressed in the form,

$$\epsilon=\epsilon_0+\alpha N, \tag{1.1}$$

where ϵ_0 is the permittivity of free space,[2] α is the polarizability of a metallic element, and N is the number of such elements per cubic meter. The conditions referred to are as follows:

(1) The metallic elements of the artificial dielectric are less than a wavelength apart.

(2) The dimensions of the metallic elements in the direction of propagation are negligible.

(3) The dimensions of these elements normal to the direction of propagation are much less than a wavelength.

An example of such a structure is an array of metallic disks supported by sheets of polystyrene. The polarizability of a disk is

$$\alpha=16\epsilon_0 a^3/3, \tag{1.2}$$

where a is the radius of a disk. Provided that the centers of the disks are less than a wavelength apart, that $a\ll\lambda$, and that the direction of propagation is normal to the plane of the disks, Eq. (1) will hold, N being the number of disks per cubic meter.

The polarization, $\mathbf{P}$, of such an artificial dielectric is given by

$$\mathbf{P}=\alpha N\mathbf{E}, \tag{1.3}$$

where $\mathbf{E}$ is the electric field strength, and hence the current will be

$$\mathbf{J}=\partial\mathbf{P}/\partial t=\alpha N(\partial\mathbf{E}/\partial t). \tag{1.4}$$

A fixed magnetic field, $\mathbf{B}_0$, will be taken in the z direction, which will also be considered the direction of propagation. There is then a Hall effect; that is, the currents in the metallic elements and the fixed magnetic field result in a secondary electric field at right angles to both the driving field and to $\mathbf{B}_0$. This field is

$$\mathbf{E}_H=R\mathbf{B}_0\times\mathbf{J}, \tag{1.5}$$

where R is the Hall coefficient[3] of the material of the metallic elements. Because of this field, there is a contribution to the dielectric displacement of an amount

$$\mathbf{D}_H=\alpha N\mathbf{E}_H=\alpha NR\mathbf{B}_0\times\mathbf{J}=\alpha^2N^2R\mathbf{B}_0\times\partial\mathbf{E}/\partial t, \tag{1.6}$$

where we have used Eq. (1.4). Hence, the total dielectric displacement, $\mathbf{D}$, is

$$\mathbf{D}=(\epsilon_0+\alpha N)\mathbf{E}+\alpha^2N^2R\mathbf{B}_0\times\partial\mathbf{E}/\partial t. \tag{1.7}$$

Eliminating the flux density $\mathbf{B}$ from the field equations in the usual manner, and substituting for $\mathbf{D}$ the value given by Eq. (1.7), we obtain

$$\nabla^2\mathbf{E}-\text{grad div}\mathbf{E}=\mu_0(\epsilon_0+\alpha N)(\partial^2\mathbf{E}/\partial t^2)+\mu_0\alpha^2N^2R\mathbf{B}_0\times\partial^3\mathbf{E}/\partial t^3, \tag{1.8}$$

which is the basic set of differential equations of this paper.[4] In general, $\text{div}\mathbf{E}\neq 0$. μ_0 is the permeability of free space.

II. FARADAY EFFECT

For an investigation of this effect it is sufficient to suppose an infinite medium (no boundary conditions on Eq. (1.8)). Setting $E_z=0$, and writing

$$p=\alpha N/\epsilon_0. \tag{2.1}$$

Equations (1.8) may be written,

$$\partial^2E/\partial z^2=(1+p/c^2)(\partial^2E/\partial t^2)+i(\epsilon_0p^2RB_0/c^2)\times\partial^3E/\partial t^3, \tag{2.2}$$

where

$$E=E_x+iE_y. \tag{2.3}$$

The presence of the last term on the right of Eq. (2.2) prevents the propagation of plane polarized waves. However, the oppositely circularly polarized waves, $\exp[i\omega(t-z/v_1)]$ and $\exp[-i\omega(t-z/v_2)]$ are found to

* Chairman, Department of Physics, University of Georgia, Atlanta, Georgia.
† Submitted to the faculty for consideration as a Ph.D. thesis.
[1] Winston E. Koch, Bell System Tech. J., January, 1948.
[2] We use MKS rationalized units throughout this paper.
[3] This Hall coefficient has the dimensions of cubic meters per coulomb, and it is numerically equal to 10^{-7} times the values listed in the International Critical Tables.
[4] See Appendix.

be solutions provided that

$$1/v_1^2=(1+p-\epsilon_0 p^2 RB_0\omega)/c^2 \tag{2.4}$$

$$1/v_2^2=(1+p+\epsilon_0 p^2 RB_0\omega)/c^2. \tag{2.5}$$

Thus, the two waves are propagated with different velocities. Combining them, we find that

$$E=A\exp(i\omega z/u)\cos\omega(t-z/v), \tag{2.6}$$

where A is an arbitrary amplitude, and where

$$1/v=(1/v_1+1/v_2)/2, \tag{2.7}$$

$$1/u=(1/v_2-1/v_1)/2. \tag{2.8}$$

From Eqs. (2.3) and (2.6),

$$E_y/E_x=\tan\omega z/u, \tag{2.9}$$

and hence ϕ, the angle of polarization, is

$$\phi=\omega z/u. \tag{2.10}$$

For a fixed z we have a linearly polarized wave, but, as z increases, the angle of polarization increases proportionally. This is precisely the Faraday effect. Just as in natural dielectrics, the effect is due to the different propagation velocities of the right- and left-handed circularly polarized waves. The cause of the difference of velocity is not, however, the same.

Practically, in the microwave region,

$$\epsilon_0 p^2 RB_0\omega \ll 1+p. \tag{2.11}$$

Hence, Eq. (2.7) gives

$$1/v^2=\mu_0(\epsilon_0+\alpha N), \tag{2.12}$$

which shows that the field propagates with nearly the same velocity that it would have in the absence of an applied magnetic field. Similarly, Eq. (2.8) yields

$$1/u=(\epsilon_0 p^2 RB_0\omega)/[2c(1+p)^{\frac{1}{2}}], \tag{2.13}$$

so that Eq. (2.10) becomes

$$\phi=(\epsilon_0 p^2 R)/[2c(1+p)^{\frac{1}{2}}]\omega^2 B_0 z. \tag{2.14}$$

The Faraday effect may be expressed by

$$\phi=V\omega^2 B_0 z, \tag{2.15}$$

where V is called Verdets' constant. Comparing Eqs. (2.14) and (2.15), we find that, for the artificial dielectric, Verdets' constant is given by

$$V=(\epsilon_0 p^2 R)/[2c(1+p)^{\frac{1}{2}}]. \tag{2.16}$$

The numerical magnitudes of these results will be discussed in the last section.

III. RESONANCE SPLITTING

The problem considered in this section is that of estimating the resonance frequencies to be expected when a cavity resonator, filled with an artificial dielectric, is subjected to a magnetic field. Formally, we have to find solutions of Eqs. (1.8) which are such that the tangential components of the electric intensity vanish at the boundaries of the resonator. This would be a difficult problem to solve rigorously because of the dissymetry introduced by the magnetic field. However, we can find formulas for the splitting provided we assume, as would be practically the case, that the effect is small.

We first assume that there are walls at only $y=0$ and $y=L$, letting the x and z directions be infinite. Then Eqs. (1.8) become

$$\partial^2 E_x/\partial y^2=\partial^2/\partial t^2[aE_x-b(\partial E_y/\partial t)], \tag{3.1}$$

$$0=\partial^2/\partial t^2[aE_y+b(\partial E_x/\partial t)], \tag{3.2}$$

$$0=\partial^2/\partial t^2(aE_z), \tag{3.3}$$

where

$$a=\mu_0(\epsilon_0+\alpha N),\quad b=\mu_0\alpha^2 N^2 RB_0. \tag{3.4}$$

Here we may set $E_z=0$, and it is easy to solve Eqs. (3.1) and (3.2) for E_x and E_y. The condition for a nontrivial periodic solution is found to be

$$\omega^4-(a^2/b^2)\omega^2+(a^2/b^2)\omega_0^2=0, \tag{3.5}$$

where ω_0 is the frequency corresponding to $\mathbf{B}_0=0$. The interesting roots of this quartic are approximately

$$\omega=\omega_0\pm(b^2\omega_0^3/2a^2), \tag{3.6}$$

so that the change in ω due to the magnetic field is only the minute second term on the right.

This situation is obviously unchanged if we take out the walls at $y=0$ and $y=L$ and introduce walls at $x=0$ and $x=L$.

Let us now consider the more interesting case of walls at $z=0$ and $z=L$, letting the x and y directions be infinite. Then Eqs. (1.8) reduce to

$$(\partial^2 E_x/\partial z^2)=\partial^2/\partial t^2[aE_x-b(\partial E_y/\partial t)] \tag{3.7}$$

$$(\partial^2 E_y/\partial z^2)=\partial^2/\partial t^2[aE_y+b(\partial/\partial t)E_x]. \tag{3.8}$$

The solution is again easy, and the condition for a nontrivial result is that ω be such as to satisfy either of the cubics

$$\omega^3\mp(a/b)\omega^2\pm(a/b)\omega_0^2=0, \tag{3.9}$$

the interesting roots of which are approximately

$$\omega=\omega_0\pm(b\omega_0^2/2a). \tag{3.10}$$

The effect shown by Eq. (3.10), while small, is very large compared with that shown by Eq. (3.6), the ratio of $b\omega_0^2/2a$ to $b^2\omega_0^3/2a$ being at least several hundred.

It is therefore reasonable that Eq. (3.10) approximates the resonance frequencies which would be observed in a rectangular or cylindrical resonator under the conditions assumed. The argument presented here corresponds to physical expectation in that the splitting is due to the effect of the magnetic field upon the velocity of the waves composing the standing wave system within the resonator, and these velocities are not seriously affected in directions normal to the field.

IV. APPLICATIONS

Two formulas which would be of use in connection with such possible applications as frequency or phase modulation, rotation of polarization, broad-banding, etc., may be obtained by minor transpositions of Eqs. (2.13) and (2.14). They are

$$\delta/z=6/\lambda^2[p^2/(1+p)^{\frac{1}{2}}]RB_0 \text{ degrees/meter} \quad (4.1)$$

where δ/z is the phase difference of the right- and left-handed circularly polarized waves, and

$$\phi/z=3/\lambda^2[p^2/(1+p)^{\frac{1}{2}}]RB_0 \text{ degrees/meter} \quad (4.2)$$

where ϕ/z is the change in the polarizing angle per unit length. Obviously, $\phi/z=\delta/2z$.

To have a particular numerical case, we assume centimeter waves, and take $B_0=0.1$ weber/m^2 (1000 gauss) with $p=8$. Then

$$\delta/z=1.28\times10^5R, \quad \phi/z=0.64\times10^5R. \quad (4.3)$$

For good metals, this gives only $\delta/z\sim10^{-5}$ degrees/meter, while for N-type germanium,[5] $\delta/z=1024$ degrees/meter and $\phi/z=512$ degrees/meter.

The phase change and rotation produced by the germanium are large enough to be encouraging. Unfortunately, the "displacement" current in the metallic elements of the artificial dielectric dissipates energy proportional to the resistivity of these elements. The resistivity of germanium is of the order of 10^6 times that of a good metal, so these losses would definitely be excessive in ordinary applications. This situation, high Hall coefficient associated with high resistivity is general among all the known materials, so that, at present, the outlook for application does not seem promising.

It will be noted, however, that the ratio of Hall coefficient to resistivity, which would be important in applications, is a quantity which varies widely for different materials. For copper, this ratio is 0.003, while for bismuth it is 0.9. This variation occurs because resistivity is mainly a function of the mean free path of conducting electrons, while the Hall effect depends on an excess or deficiency of electrons in the crystal lattice. Furthermore, the Hall coefficient is very sensitive to very small amounts of impurities.[6] In view of these complications it would seem premature to say that no material with a sufficiently high ratio of Hall coefficient to resistivity to be useful could be found.

[5] G. L. Pearson [Rev. Sci. Instr. **19**, 263 (1948)] gives $R=8\times10^{-3}$.

APPENDIX

While this paper was in manuscript form, some questions arose as to the effect of interaction on the derivation of Eq. (1.8). The objection arises because the currents in the conducting elements, excited by the impressed field, are themselves sources of a field.

Strictly speaking, this is an objection to our starting point, Eq. (1.1), which is derived by neglecting the induced fields. Perhaps no aspect of the theory of artificial dielectrics has been more exhaustively analyzed than this.[7] Exact formulas, which approximate Eq. (1.1) in the limit indicated by our conditions (1) and (3), have been worked out for the cases where the conducting elements are spheres, cylinders, strips, squares, and disks. Generally speaking, Eq. (1.1) is good to within about 10 percent for dielectric constants up to 1.5.

The point might be made that, having neglected the conducting elements as a source of field in writing down Eq. (1.1), consistency might seem to require that we also neglect the small Hall field which we needed to get a Faraday effect. However, the Hall field is effective in changing the direction of polarization in a decisive way, while the induced field is not.

[6] G. L. Pearson, Rev. Sci. Instr. **19**, 263 (1948). Lark-Horowitz, Middleton, Miller, and Walerstein, Phys. Rev. **69**, 258(A) (1946).

[7] L. Lewis, J. Inst. Elec. Engrs. (January 1947). S. Pasternack, "Microwave lenses" Final Report under Contract No. W 28-099-ac-112, University of Pennsylvania, July, 1947. S. B. Cohn, J. Appl. Phys. **20**, 257–262 (1949). J. Brown, Proc. Inst. Elec. Engrs. 97 (III), 1950. S. B. Cohn, J. Appl. Phys. **21**, 674–680 (1950).

General Theory of Symmetric Biconical Antennas*

S. A. SCHELKUNOFF
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received April 25, 1951)

This paper presents the input admittance of a symmetric biconical antenna of an arbitrary angle as the limit of a certain sequence of functions. The first term of this sequence approaches the exact expressions for the input admittance as the cone angle approaches either zero or 90°. For this reason our conjecture is that this term represents a good first approximation for all angles.

IN this paper we consider antennas formed by two equal, coaxial, and perfectly conducting cones whose apices are infinitely close, Fig. 1. At the ends of the cones we assume perfectly conducting spherical segments of radius l equal to the generators of the cones. Our problem is to find that particular solution of the following differential equations:

$$(\partial/\partial r)(rH_\varphi)=-i\omega\epsilon rE_\theta,\quad (\partial/\partial\theta)(\sin\theta H_\varphi)=i\omega\epsilon r\sin\theta E_r,$$
$$(\partial/\partial r)(rE_\theta)-\partial E_r/\partial\theta=-i\omega\mu rH_\varphi, \tag{1}$$

which satisfies the following boundary conditions:

(A) $E_r(r,\psi)=E_r(r,\pi-\psi)=0,\ 0<r<l$;

(B) $E_\theta(l,\theta)=0,\ 0\leq\theta<\psi$ and $\pi-\psi<\theta\leq\pi$;

(C) rE_θ, rH_φ vary as $\exp(-i\beta r)$ as r approaches infinity.

(D) rE_θ, rH_φ approach finite and nonvanishing limits as r approaches zero;

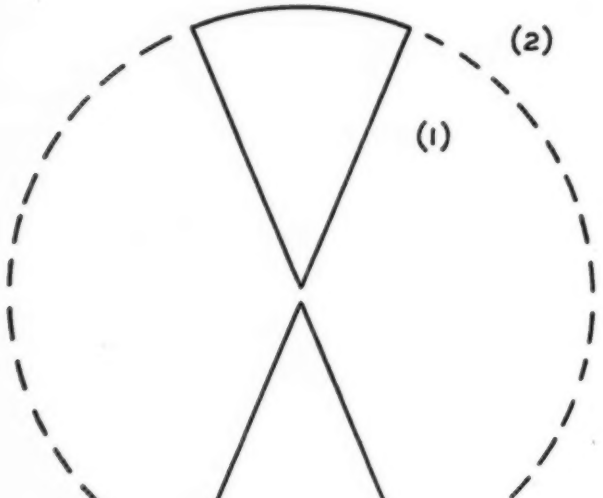


FIG. 1. A symmetric biconical antenna and a division of space into two regions.

(E) $\lim\int_{\psi}^{\pi-\psi} rE_\theta(r,\theta)d\theta$ as r approaches zero is given; this limit equals the electromotive force driving the antenna.

In addition, we require that the solution have no singularities outside the double cone.

In particular, we wish to determine the "input admittance" of the biconical antenna, which is defined as the ratio

$$Y_i=I(0)/V(0), \tag{2}$$

where

$$I(0)=\lim 2\pi r\sin\psi H_\varphi(r,\psi),\ r\to 0,$$
$$V(0)=\lim\int_{\psi}^{\pi-\psi} rE_\theta(r,\theta)d\theta,\ r\to 0. \tag{3}$$

The method of solution is based on separating the variables in Eq. (1) and subdividing the region external to the double cone into two subregions (1) and (2) by a "boundary sphere" of radius l centered at the apices of the cones. In region (1) we shall write the most general solution satisfying the boundary conditions (*A*), (*D*), (*E*); in region (2) we shall write the most general solution satisfying the boundary conditions (*B*), (*C*); then we shall seek that particular solution which is continuous across the boundary sphere. We add, therefore, the following boundary condition:

(*F*) $$E_\theta^{(1)}(l,\theta)=E_\theta^{(2)}(l,\theta),\ \psi<\theta<\pi-\psi;$$
$$H_\varphi^{(1)}(l,\theta)=H_\varphi^{(2)}(l,\theta),\ \psi<\theta<\pi-\psi.$$

The continuity of E_r follows from the continuity of H_φ and need not be stated explicitly.

The special cases in which ψ is very small or nearly equal to $\pi/2$ have been considered by the author.[1] The limiting case of $\psi=\pi/2$ has also been considered by Stratton and Chu.[2] Approximate solutions in the intermediate range of ψ have been given by Smith[3] and by Tai.[4] Here we shall give the general solution in the form of an infinite series of which the first term reduces to the exact limiting forms as ψ approaches either zero or $\pi/2$. This term expresses a first approximation to the solution for all cone angles.

* Presented at the International Congress of Mathematicians at Harvard University, Cambridge, Massachusetts, August 30 to September 6, 1950.

[1] S. A. Schelkunoff, Proc. Inst. Radio Engrs. **29**, 493–521 (1941), U. S. Patent 2,235,506, issued March 18, 1941.
[2] J. A. Stratton and L. J. Chu, J. Appl. Phys. **12**, 236–240 (1941).
[3] P. D. P. Smith, J. Appl. Phys. **19**, 11–23 (1948).
[4] C. T. Tai, J. Appl. Phys. **20**, 1076–1084 (1949).

In either region the solution may be written in the following form:

$$\omega\epsilon r^2 E_r = i\sum_n n(n+1)R_n(r)\Theta_n(\theta),$$
$$\omega\epsilon r E_\theta = i\sum_n R_n'(r)\Theta_n'(\theta), \quad rH_\varphi = \sum_n R_n(r)\Theta_n'(\theta), \tag{4}$$

where in each region n is a typical member of an appropriate set of separation constants and $R_n(r)$, $\Theta_n(\theta)$ are solutions of

$$\sin\theta(d^2\Theta_n/d\theta^2)+\cos\theta(d\Theta_n/d\theta)+n(n+1)\sin\theta\Theta_n=0, \tag{5}$$

$$(d^2R_n/dr^2)=[-\beta^2+(n(n+1)/r^2)]R_n, \beta^2=\omega^2\mu\epsilon. \tag{6}$$

Without loss of generality we may assume that n is non-negative.

In region (1) we must satisfy (A) for all values of r; hence, (A) must be satisfied by the individual terms of the series for E_r. Consequently, n must either vanish or be a root of

$$\Theta_n(\psi)=\Theta_n(\pi-\psi)=0. \tag{7}$$

If $n=0$, the general solution of (5) is

$$\Theta_0(\theta)=a_0\log\tan\tfrac{1}{2}\theta+b_0. \tag{8}$$

Without loss of generality in (4) we may assume $b_0=0$. The corresponding solution of (6) is

$$R_0(r)=A_0\sin\beta r+B_0\cos\beta r. \tag{9}$$

If $n\neq 0$ and is not an integer the general solution of (5) and the corresponding solution of (6) are

$$\Theta_n(\theta)=a_nP_n(\cos\theta)+b_nP_n(-\cos\theta), \tag{10}$$

$$R_n(r)=A_nJn_n(\beta r)+B_nNn_n(\beta r), \tag{11}$$

where $P_n(\cos\theta)$, $P_n(-\cos\theta)$ are legendre functions and $Jn_n(\beta r)$, $Nn_n(\beta r)$ are normalized bessel functions defined by

$$Jn_n(x)=(\pi x/2)^{\frac{1}{2}}J_{n+\frac{1}{2}}(x),\ Nn_n(x)=(\pi x/2)^{\frac{1}{2}}N_{n+\frac{1}{2}}(x). \tag{12}$$

As $x\to 0$, the second of these functions varies as x^{-n}; hence, to satisfy the condition D we must let $B_n=0$. To satisfy Eq. (7) we find that Θ must be either an even or an odd Legendre function of $\cos\theta$,

$$L_n(\cos\theta)=\tfrac{1}{2}[P_n(\cos\theta)+P_n(-\cos\theta)]$$
$$M_n(\cos\theta)=\tfrac{1}{2}[P_n(\cos\theta)-P_n(-\cos\theta)]. \tag{13}$$

The separation constant n must be a root of

$$P_n(\cos\psi)=-P_n(-\cos\psi) \tag{14}$$

in the first case, and a root of

$$P_n(\cos\psi)=P_n(-\cos\psi) \tag{15}$$

in the second. We note that the function defined by (8), with $b_0=0$, is an odd function of $\cos\theta$.

If n is an integer, Eq. (10) does not represent the general solution. In this case we express it as a linear combination of $P_n(\cos\theta)$ and $Q_n(\cos\theta)$. For integral values of n one of these functions is an even function of $\cos\theta$ and the other is an odd function.

In region (2) there are no solutions of (5) which are finite for all values of θ except when n is an integer in which case the finite solutions are given by

$$\Theta_n(\theta)=a_nP_n(\cos\theta). \tag{16}$$

These functions are even or odd functions of $\cos\theta$ depending on whether n is even or odd. The corresponding $R_n(r)$ is given by (11) in which, in order to satisfy the condition (C), we must choose

$$B_n=-iA_n.$$

Hence,

$$R_n(r)=A_n[Jn_n(\beta r)-iNn_n(\beta r)]. \tag{17}$$

We have now determined the solutions in regions (1) and (2) which satisfy the boundary conditions (A), (B), (C), (D). Of these we have to select one which satisfies (E) and (F). The latter condition must be satisfied separately by the even and odd parts of the total solutions. For all even functions which contribute to the solution in region (1), n is greater than zero. While b_0 is an even function of $\cos\theta$ in Eq. (8) corresponding to $n=0$, it does not contribute anything to the field given by (4). We have seen that R_n in region (1) is given by (11) where $B_n=0$. As r approaches zero, the first term approaches zero as r^{n+1}; hence the various terms contributing to rE_θ vary as r^n. Consequently, the even part of the solution contributes nothing to the last integral in the set (3), that is, the even part of the solution is independent of the impressed electromotive force and does not represent a response to it. It can be shown that under the conditions stated at the beginning of this paper the even part must actually vanish; but we shall not give this matter further attention and seek the odd solution to our problem.

In view of the foregoing considerations we can write the following expressions for region (1):

$$rH_\varphi=(I_0(r)/2\pi\sin\theta)$$
$$+\sum_n a_n(Jn_n(\beta r)/Jn_n(\beta l))(d/d\theta)M_n(\cos\theta), \tag{18}$$

$$rE_\theta=(\eta V(r)/2\pi K\sin\theta)$$
$$+i\eta\sum_n a_n(Jn_n'(\beta r)/Jn_n(\beta l))(d/d\theta)M_n(\cos\theta), \tag{19}$$

where

$$\eta=(\mu/\epsilon)^{\frac{1}{2}},\ K=(\eta/\pi)\log\cot\tfrac{1}{2}\psi, \tag{20}$$

$$V(r)=V(l)[\cos\beta(l-r)+iKY_l\sin\beta(l-r)], \tag{21}$$

$$KI_0(r)=V(l)[i\sin\beta(l-r)+KY_l\cos\beta(l-r)], \tag{22}$$

and the summation is extended over all the roots of (15). The first terms in (18) and (19) arise from the characteristic functions given by (8) and (9). The particular form adopted for the first terms is suggested by physical considerations and happens to be convenient for expressing the input admittance (2). In electric "transmission lines" the voltage and current are distributed in accordance with (9) where r is the

distance along the line and A_0, B_0 are constants depending on the end conditions. In such lines the voltage and current may be expressed as in (21) and (22) where $V(l)$ is the voltage at $r=l$, Y_t is the "terminating admittance," and K is the "characteristic impedance." The particular value of K given by (20) is the value we obtain for the ratio $V(0)/I(0)$ when the cones are extended to infinity so that all the a_n's must vanish.†

In region (2) we have

$$rH_\varphi=\sum_k b_k(Hn_k(\beta r)/Hn_k(\beta l))(d/d\theta)P_k(\cos\theta),$$

$$H_{nk}(\beta r)=Jn_k(\beta r)-iNn_k(\beta r), \quad (23)$$

$$rE_\theta=i\eta\sum_k b_k(Hn_k'(\beta r)/Hn_k(\beta l))(d/d\theta)P_k(\cos\theta),$$

$$k=1, 3, 5, \cdots \quad (24)$$

We now obtain $I(0)$ and $V(0)$ defined by (3) from Eqs. (18) to (22):

$$I(0)=I_0(0)=K^{-1}V(l)(KY_t\cos\beta l+i\sin\beta l),$$

$$V(0)=V(l)(\cos\beta l+iKY_t\sin\beta l). \quad (25)$$

Therefore, the required normalized input admittance (2) is

$$KY_i=\frac{KY_t\cos\beta l+i\sin\beta l}{\cos\beta l+iKY_t\sin\beta l}. \quad (26)$$

Letting $r=l$ in (18) and integrating from $\theta=\psi$ to $\theta=\pi-\psi$, we find

$$KY_t=\eta/V(l)\int_\psi^{\pi-\psi} lH_\varphi(l,\theta)d\theta. \quad (27)$$

In view of the boundary condition (F) we can substitute for lH_φ the series (23); thus

$$KY_t=-(2\eta/V(l))\sum_k b_kP_k(\cos\psi). \quad (28)$$

We now let $r=l$ in (24), use the orthogonal properties of the derivatives of the Legendre functions and the boundary condition (B) to obtain

$$b_k=(Hn_k(\beta l)/i\eta N_k{}^eHn_k'(\beta l))\int_\psi^{\pi-\psi} lE_\theta(l,\theta)\times\sin\theta(d/d\theta)P_k(\cos\theta)d\theta, \quad (29)$$

where $N_k{}^e$ is the normalizing factor for the exterior Legendre functions,

$$N_k{}^e=\int_0^{2\pi}[(d/d\theta)P_k(\cos\theta)]^2\sin\theta d\theta=2k(k+1)/2k+1. \quad (30)$$

† For rH_φ and rE_θ must then satisfy the condition (C).

Similarly from (18) we have,

$$a_n=(N_n{}^i)^{-1}\int_\psi^{\pi-\psi} lH_\varphi(l,\theta)\sin\theta(d/d\theta)M_n(\cos\theta)d\theta, \quad (31)$$

where

$$N_n{}^i=\int_\psi^{\pi-\psi}[(d/d\theta)M_n(\cos\theta)]^2\sin\theta d\theta$$

$$=-(2n(n+1)/2n+1)\sin\psi(\partial M_n(\cos\psi)/\partial\psi)\times(\partial M_n(\cos\psi)/\partial n)$$

$$=-(2n(n+1)/2n+1)\sin\psi((\partial M_n/\partial\psi)^2)(d\psi/dn), \quad (32)$$

where ψ as a function of n is given by (15).

Using the boundary condition (F) and substituting from (19) in (29) and from (23) in (31), we express the b's in terms of the a's and the a's in terms of the b's. Eliminating the a's, we obtain an infinite system of linear equations in the b's. Thus

$$(N_k{}^eZ_k{}^e+Z_{kk})b_k=-(\eta/\pi K)V(l)P_k(\cos\psi)-\sum_\alpha{}'Z_{k\alpha}b_\alpha, \quad (33)$$

where the prime signifies the omission of the term corresponding to $\alpha=k$ and

$$Z_k{}^e=(i\eta Hn_k'(\beta l)/Hn_k(\beta l)),$$

$$Z_n{}^i=(\eta Jn_n'(\beta l)/iJn_n(\beta l)) \quad (34)$$

$$Z_{k\alpha}=-\sum_n\times\frac{(2n+1)k(k+1)\alpha(\alpha+1)\sin\psi P_k(\cos\psi)P_\alpha(\cos\psi)}{2n(n+1)(k-n)(\alpha-n)(k+n+1)(\alpha+n+1)}\times\frac{dn}{d\psi}Z_n{}^i. \quad (35)$$

Equation (33) may be solved by successive approximations. We write

$$b_k=b_k{}^{(0)}+b_k{}^{(1)}+b_k{}^{(2)}+\cdots \quad (36)$$

where

$$b_k{}^{(0)}=-\frac{\eta V(l)P_k(\cos\psi)}{\pi K(N_k{}^eZ_k{}^e+Z_{kk})},$$

$$b_k{}^{(m+1)}=-\frac{\sum_\alpha{}'Z_{k\alpha}b_\alpha{}^{(m)}}{N_k{}^eZ_k{}^e+Z_{kk}}. \quad (37)$$

From the b's we obtain kY_t and then the input admittance Y_i.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

The Transit Time Effect in a Cylindrical Diode

WAY DONG WOO
Computation Laboratory, Harvard University, Cambridge, Massachusetts
(Received May 2, 1951)

The impedance offered to a small superposed alternating current by a space-charge limited cylindrical diode is expressed in terms of the variational plate conductance, the transit time angle, and a group of constants which are functions of the ratio of the plate to cathode radii. The result covers the range of transit-time angle up to π-radian and of the ratio of the plate to cathode radii from unity up to 100.

I. INTRODUCTION

FOR very low frequencies, when the transit time of electrons from cathode to plate is negligible compared with the period of a small superposed alternating current, the impedance of a diode to such an alternating current can be considered as the variational plate resistance shunted by the interelectrode capacity. When the frequency is increased, it no longer holds true. Benham,[1] Müller,[2] and Llewellyn[3] investigated the transit time effect in the impedance offered by a parallel plane diode to a small alternating current superposed on a large steady current. For a cylindrical diode, the problem is much more difficult because the differential equations involved in this problem do not give solutions in simple closed functions. In this work, numerical methods are used. The impedance is expressed as a power series of transit time angle, which is equal to 2π times the ratio of the transit time to the period of the applied alternating current.

II. THE DIFFERENTIAL EQUATION OF MOTION OF ELECTRONS

The following assumptions are made:

(1) Complete cylindrical symmetry and no circumferential variations in all quantities concerned are assumed. Cathode is the inner electrode.

(2) Cathode is completely space-charge limited, so that the initial velocity and acceleration are zero.

(3) The interelectrode spacing is negligible compared with the wavelength of the applied alternating current, so that the potential difference between cathode and plate is the line integral of the electric field.

(4) The magnetic field due to moving electrons can be neglected.

(5) The superposed alternating current is small compared with the steady direct current. The electrons always move toward plate and never overtake one another.

The first four assumptions are satisfied in all practical cases, and the last one limits the scope of the present investigation to the first-order effect of the superposed current.

Let r_c = radius of cathode in meter, r_p = radius of the plate in meter, $r_p > r_c$, t = time of observation in sec, t_a = time at which an electron leaves the cathode, and $r = r(t, t_a)$ = radial coordinate in meter at t of the electrons which leave the cathode at t_a. Then it follows that

$$mr'' = m\partial^2 r/\partial t^2 = -eE, \tag{1}$$

where m = mass of electron in kg, e = numerical value of charge of electron in coulombs, $e/m = 1.76\times10^{11}$ coulombs/kg, and $E = E(r, t)$ = radial electric field at coordinate r and time t.

Let ϵ_0 = dielectric constant of vacuum $= 1/36\pi\times10^{-9}$ coulomb/volt meters, ρ = charge density in coulombs/meter^3, and v = velocity of electron in meter/sec. Then, from electromagnetic wave equations, one obtains

$$\rho = \epsilon_0 \operatorname{div} E = (\epsilon_0/r)\partial(rE)/\partial r \tag{2}$$

and

$$\operatorname{div}(\rho v + \epsilon_0 \partial E/\partial t) = (1/r)\partial[r\rho v + \epsilon_0 r \partial E/\partial t]/\partial r = 0.$$

Thus $r(\rho v + \epsilon_0 \partial E/\partial t)$ is a function of time only, so that

$$2\pi r(\rho v + \epsilon_0 \partial E/\partial t) = 2\pi r_c(\rho v + \epsilon_0 \partial E/\partial t)_{r=r_c} = -J(t). \tag{3}$$

Obviously $-J$ is the internal current from cathode to plate per unit length of the diode. Substituting Eq. (2) into (3), one obtains

$$-J(t) = 2\pi\epsilon_0 r\left[\frac{v}{r}\frac{\partial}{\partial r}(rE) + \frac{\partial E}{\partial t}\right]$$

$$= 2\pi\epsilon_0\left[v\frac{\partial}{\partial r}(rE) + \frac{\partial}{\partial t}(rE)\right] = 2\pi\epsilon_0 d(rE)/dt. \tag{4}$$

The total derivative $d(rE)/dt$ denotes the time rate of rE as seen by an observer moving with the electron. Substituting Eq. (1) into Eq. (5) and rearranging terms, one obtains

$$d(rr'')/dt = Je/2\pi m\epsilon_0. \tag{5}$$

This is the equation of motion of electrons. The initial conditions are:

$$\text{at } t = t_a,\ r = r_c,\ r' = \partial r(t, t_a)/\partial t = 0,$$
$$r'' = \partial^2 r(t, t_a)/\partial t^2 = 0. \tag{6}$$

III. SEPARATION INTO DC AND AC COMPONENTS AND REDUCTION INTO DIMENSIONLESS QUANTITIES

The current density $J_0(t)$ is composed of the large steady part J_0 which is a constant, and the small varia-

[1] W. E. Benham, Phil. Mag. (March, 1928); *ibid.*, Suppl. 11 (February, 1931).
[2] J. Müller, Hochfrequenztech und Elektroakustik, 41 (May, 1933).
[3] F. B. Llewellyn, *Electron Inertia Effect* (Cambridge University Press, London, 1941).

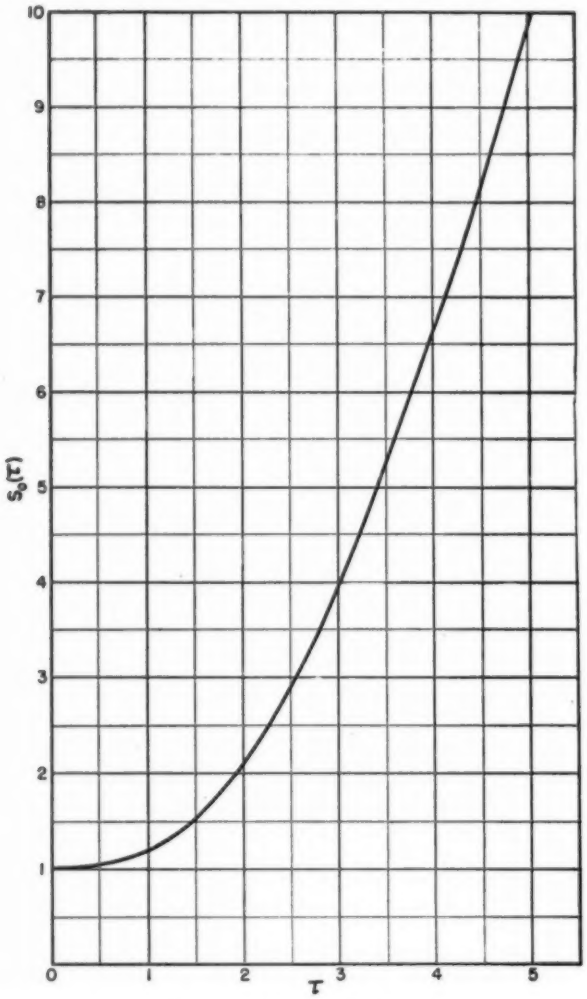


FIG. 1. $S_0(\tau)$ *vs* τ, $0 \leqslant \tau \leqslant 5$.

tional part $j(t)$:

$$J(t)=J_0+j(t). \tag{7}$$

From Eq. (1), one obtains

$$d(rr'')/dt=[J_0+j(t)]e/2\pi m\epsilon_0. \tag{8}$$

Let r_0 be the solution of Eq. (8) when $j(t)\equiv 0$; then

$$\left.\begin{aligned} &d(r_0r_0'')/dt=J_0e/2\pi m\epsilon_0, \\ &r_0=r_c, \quad r_0'=0, \quad r_0''=0, \quad \text{at} \quad t=t_a, \end{aligned}\right\} \tag{9}$$

where r_0 is the radial coordinate of an electron when no alternating current is applied.

Let

$$r=r_0+r_1. \tag{10}$$

Then r_1 is the displacement due to the superimposed current. Substituting this into Eq. (5), one obtains

$$d(r_0r_1''+r_0''r_1+r_1r_1'')/dt=(e/2\pi m\epsilon_0)j(t). \tag{11}$$

Since $j(t)$ is assumed to be very small compared with J_0, one has $r_1 \ll r_0$, so that the last term on the left side of Eq. (11) may be neglected. Thus, one obtains a linear differential equation for r_1, which is the variational component of the motion of the electrons

$$d(r_0r_1''+r_0''r_1)/dt=(e/2\pi m\epsilon_0)j(t). \tag{12}$$

For the convenience of computation, the variables r_0, r_1, t, J, and j are converted into dimensionless quantities. Let

$$\left.\begin{aligned} &S=r/r_c, \quad S_0=r_0/r_c, \quad S_1=r_1/r_c, \\ &\theta=t/\alpha, \quad \theta_a=t_a/\alpha, \\ &\alpha=[2\pi\epsilon_0 mr_c^2/J_0e]^{\frac{1}{2}}=6.81\times 10^{-8}[r_c^2/J_0]^{\frac{1}{2}}, \\ &k(\theta)=j(t)/J_0=j(\alpha\theta)/J_0; \end{aligned}\right\} \tag{13}$$

then

$$S(\theta, \theta_a)=S_0(\theta, \theta_a)+S_1(\theta, \theta_a), \tag{14}$$

$$d(S_0S_0'')/d\theta=1, \tag{15}$$

$$d(S_0S_1''+S_0''S_1)/d\theta=k(\theta), \tag{16}$$

$$S_0=1, \quad S_0'=0, \quad S_0''=0, \quad S_1=0, \quad S_1'=0, \quad S_1''=0 \quad \text{at} \quad \theta=\theta_a. \tag{17}$$

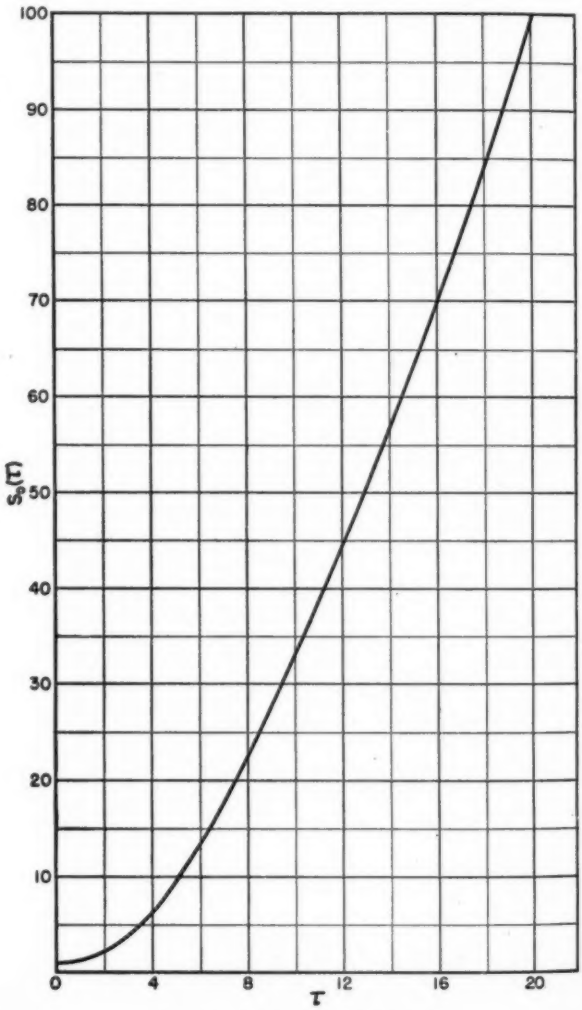


FIG. 2. $S_0(\tau)$ *vs* τ, $0 \leqslant \tau \leqslant 20$.

IV. SOLUTION FOR STEADY STATE

In the steady state the position of an electron depends only on how long it has left the cathode. Thus, $r_0(t, t_a)$ is a function of $t-t_a$, and $S_0(\theta, \theta_a)$ is a function of $\theta-\theta_a$. Defining

$$\tau=\theta-\theta_a, \tag{18}$$

one obtains

$$\left.\begin{aligned} &S_0(\theta, \theta_a)=S_0(\theta-\theta_a)=S_0(\tau),\\ &d(S_0S_0'')/d\tau=1,\\ &S_0=1, \quad S_0'=0, \quad S_0''=0 \quad \text{at} \quad \tau=0. \end{aligned}\right\} \tag{19}$$

Integrating Eq. (19), one has

$$S_0S_0''=\tau. \tag{20}$$

This equation was first solved for small value of $\tau(\tau\leqslant 1)$ by Aiken.[4] The derivatives at $\tau=0$ are obtained by repeated differentiation of the differential equation, and are then used in the Taylor's expansion at $\tau=0$. For $\tau\leqslant 1$,

$$S_0=1+\frac{\tau^3}{6}-\frac{\tau^6}{180}+\frac{\tau^9}{2160}-\frac{\tau^{12}}{19{,}008}+\cdots. \tag{21}$$

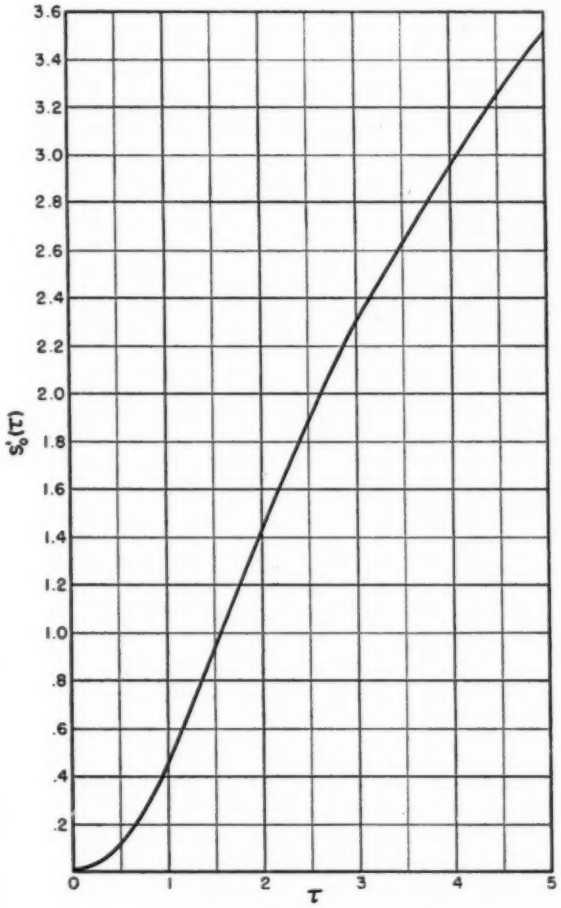


FIG. 3. $S_0'(\tau)$ *vs* τ, $0\leqslant\tau\leqslant 5$.

[4] H. Aiken, "Theory of space charge conduction," unpublished thesis, Harvard University, 1939.

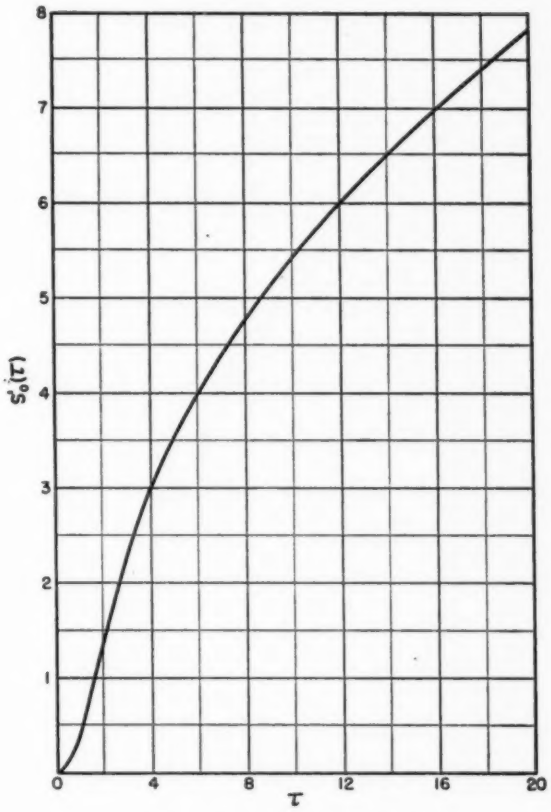


FIG. 4. $S_0'(\tau)$ *vs* τ, $0\leqslant\tau\leqslant 20$.

The numerical values of S_0 at $\tau=0$, 0.25, 0.5, 0.75, and 1.0 are found from the last equation; and they serve as the starting values for the numerical integration[5] of the differential equation (20) for larger range of τ. The results are plotted in Fig. 1 and Fig. 2. $S_0'(\tau)$ is obtained by numerical differentiation of S_0, and the result is plotted in Fig. 3 and Fig. 4.

The voltage drop across the plate and cathode is obtained by integrating the electric field from cathode to plate:

$$V_0=V_p-V_c=-\int_{\tau=0}^{\tau=\tau_p}Edr=(m/e)\int_{\tau=0}^{\tau=\tau_p}r''dr$$

$$=(m/e)\int_{\tau=0}^{\tau=\tau_p}r''r'd\tau=\tfrac{1}{2}(m/e)(r')^2\Big]_{\tau=0}^{\tau=\tau_p}$$

$$=\tfrac{1}{2}(m/e)(r_0')_p^2=\tfrac{1}{2}(m/e)(r_c^2/\alpha^2)(S_0')_p^2$$

$$=\tfrac{1}{2}(m/e)r_c^2\left[\frac{2\pi\epsilon_0 m r_c^2}{J_0^4}\right]^{-\frac{1}{3}}(S_0')_p^2, \tag{22}$$

$$R_0=V_0/J_0=\tfrac{1}{2}(m/e)(r_c^2/J_0\alpha^2)[S_0']_p^2. \tag{22a}$$

By rearranging terms of Eq. (22), the well-known three-

[5] For example, see W. E. Milne, *Numerical Calculus* (Princeton University Press, Princeton, New Jersey, 1949), p. 140.

half-power law is obtained:

$$J_0=4\pi\epsilon_0(2e/m)^{\frac{1}{2}}[V_0^{\frac{1}{2}}/r_c(S_0')_p^{\,3}]$$
$$=2.33[V_0^{\frac{1}{2}}/r_c(S_0')_p^{\,3}]\times10^{-6}, \quad (23)$$

J_0 being in amperes per meter, V_0 and r_c in volts and meters, respectively.

Langmuir and Blodgett[6] derived a current-voltage relation for cylindrical diodes. Converted into the mks unit, it is

$$J_0=(8\pi\epsilon_0/9)(2e/m)^{\frac{1}{2}}V_0^{\frac{1}{2}}/r_pB^2,$$

where B is a rapidly converging series of $\ln(r_p/r_c)$. The relation between $(S_0')_p$ in this work and B in Langmuir and Blodgett's work is then

$$(S_0')_p=[(9/2)S_pB^2]^{\frac{1}{2}}.$$

These relations are checked for many values of S_0 and are found to be true.

The variational plate conductance is obtained by differentiating Eq. (23),

$$g=\partial J_0/\partial V_0=\tfrac{3}{2}\cdot4\pi\epsilon_0(2e/m)^{\frac{1}{2}}V_0^{\frac{1}{2}}/r_c(S_0')_p^{\,3}=\tfrac{3}{2}J_0/V_0$$
$$=3.495[V_0^{\frac{1}{2}}/r_c(S_0')_p^{\,3}]\times10^{-6}\ \text{mho/meter}. \quad (24)$$

The variational resistance is, of course, two-thirds of the resistance to the direct current.

V. MOTION OF ELECTRONS WITH SUPERPOSED ALTERNATING CURRENT

For sinusoidal variation of the superposed current, one has

$$\left.\begin{aligned} &j(t)=j_1\epsilon^{i\omega t},\\ &k(\theta)=j(t)/J_0=K\epsilon^{i\omega t}=K\epsilon^{i\omega\alpha\theta}=K\epsilon^{p\theta},\\ &p=i\omega\alpha, \quad K=j_1/J_0. \end{aligned}\right\} \quad (25)$$

When this is substituted into Eq. (16), the differential equation for the variational component of motion of electron becomes

$$d(S_0S_1''+S_0''S_1)/d\theta=K\epsilon^{p\theta}. \quad (26)$$

Integrating Eq. (26), one obtains

$$(S_0S_1''+S_0''S_1)\Big]_{\theta-\theta_a}^{\theta-\theta}=\int_{\theta_a}^{\theta}K\epsilon^{p\theta}d\theta=(K/p)(\epsilon^{p\theta}-\epsilon^{p\theta_a}).$$

At $\theta=\theta_a$, S_1, S_1', S_1'', S_0', S_0'' are all zero by Eq. (17), so the last equation becomes

$$S_0S_1''+S_0''S_1=(K/p)(\epsilon^{p\theta}-\epsilon^{p\theta_a}). \quad (27)$$

Let a solution of the following form be tried:

$$S_1=(K/p)[\phi(\tau)\epsilon^{p\theta}+S_0'(\tau)(\epsilon^{p\theta}-\epsilon^{p\theta_a})]. \quad (28)$$

Then, one has

$$S_1'=\partial S_1/\partial\theta=(K/p)[(\phi'+p\phi)\epsilon^{p\theta}$$
$$+(S_0''(\tau)+pS_0'(\tau))\epsilon^{p\theta}-S_0''(\tau)\epsilon^{p\theta_a}],$$

$$S_1''=\partial^2S_1/\partial\theta^2=(K/p)[(\phi''+2p\phi'+p^2\phi)\epsilon^{p\theta}+(S_0'''(\tau)$$
$$+2pS_0''(\tau)+p^2S_0'(\tau))\epsilon^{p\theta}-S_0'''(\tau)\epsilon^{p\theta_a}].$$

If one substitutes these into Eq. (27) and uses the relation that $S_0S_0''=\tau$ and $S_0S_0'''+S_0'S_0''=(S_0S_0'')'=1$, the terms that have $\epsilon^{p\theta_a}$ as a factor cancel out, while those that have $\epsilon^{p\theta}$ as a factor satisfy the following relation:

$$S_0(\phi''+2p\phi'+p^2\phi)+S_0''\phi+2p\tau+p^2S_0S_0'=0. \quad (29)$$

In order that the initial conditions $S_1=S_1'=S_1''=0$ at $\theta=\theta_a$ be satisfied, $\phi(\tau)$, $\phi'(\tau)$ must be equal to zero at $\tau=0$. With these initial conditions on ϕ and ϕ', ϕ'' is equal to zero by Eq. (28), so that the condition $S_1''=0$ at $\theta=\theta_a$ is automatically satisfied:

$$\phi(\tau)=0, \quad \phi'(\tau)=0 \quad \text{at} \quad \tau=0. \quad (29a)$$

The function $\phi(\tau)$ involves a parameter p. When p is small, ϕ can be expanded in a power series in p. The leading term is easily verified as the first power of p:

$$\phi=C_0p+C_1p^2+C_2p^3+C_3p^4+\cdots. \quad (30)$$

Substituting Eq. (30) and its derivatives into Eq. (29), and equating coefficients of equal powers of p, one obtains

$$S_0C_0''+S_0''C_0+2\tau=0, \quad (31)$$

$$\left.\begin{aligned} &S_0C_1''+S_0''C_1+2S_0C_0'+S_0S_0'=0,\\ &S_0C_2''+S_0''C_2+2S_0C_1'+S_0C_0=0,\\ &S_0C_3''+S_0''C_3+2S_0C_2'+S_0C_1=0,\\ &S_0C_4''+S_0''C_4+2S_0C_3'+S_0C_2=0,\\ &\text{- - - - - - - - - - - - - - - - - -} \end{aligned}\right\} \quad (32)$$

The initial conditions are

$$C_0, C_1, C_2, \cdots C_0', C_1', C_2', \cdots=0 \quad \text{at} \quad \tau=0. \quad (33)$$

It is easily verified that the differential equation (31) is satisfied by $-\frac{2}{3}S_0'\tau$, for

$$\left.\begin{aligned} &C_0=-\tfrac{2}{3}S_0'\tau,\\ &C_0'=-\tfrac{2}{3}(S_0'+S_0''\tau), \quad C_0''=-\tfrac{2}{3}(2S_0''+S_0'''\tau), \end{aligned}\right\} \quad (34)$$

and, $C_0'(0)=C_0(0)=0$, satisfying the initial conditions:

$$S_0C_0''+S_0''C_0$$
$$=-\tfrac{2}{3}[2S_0''S_0+S_0S_0'''\tau+S_0''S_0'\tau]$$
$$=-\tfrac{2}{3}[2S_0S_0''+\tau(S_0S_0'+S_0'S_0'')]$$
$$=-\tfrac{2}{3}[2\tau+\tau(S_0S_0'')']=-2\tau, \text{ satisfying Eq. (31).}$$

C_0 can thus be calculated directly from the steady-state solution. The other terms C_1, C_2, C_3, etc., are then solved numerically from the recurrent differential equations.

VI. THE IMPEDANCE TO SMALL SUPERPOSED ALTERNATING CURRENT

The potential difference, at time t, between plate and cathode is obtained by integrating the electric field

[6] I. Langmuir and K. Blodgett, Phys. Rev. 22, 347 (1923).

from r_c to r_p,

$$V(t)=V_p-V_c=-\int_{R=r_c}^{R=r_p}E(R,t)dR$$

$$=(m/e)\int_{R=r_c}^{R=r_p}A(R,t)dR, \quad (35)$$

where R is an auxiliary variable of integration, and $A(R,t)$ denotes the acceleration of electrons which are at distance R from axis. In the previous section, the position of an electron is expressed as a function of t and its starting time t_a. By eliminating t_a from this function and its second derivative, the acceleration can be expressed as a function of t and R. However, this is a very difficult and cumbersome process. To circumvent this difficulty, new auxiliary variables are introduced. Let T and λ be defined as follows:

$$\left.\begin{aligned}T&=\alpha\lambda,\\ R&=r_0(T)=r_cS_0(\lambda).\end{aligned}\right\} \quad (36)$$

Physically, T is the time an electron takes to reach R from cathode when there is no superposed current. The relation between T, t, t_a, is then

$$R=r_0(T)=r_0(t-t_a)+r_1(t,t_a). \quad (37)$$

By the assumption that $r_1\ll r_0$, T is almost equal to $t-t_a$, so that $T-(t-t_a)$ is a very small quantity. Expanding $r_0(t,t_a)$ and $r_1(t,t_a)$ as a Taylor series at t, $t-T$, one obtains

$$\begin{aligned}r_1(t,t_a)&=r_0(T)-r_0(t-t_a)\\&=r_0'(T)[T-(t-t_a)]+\text{higher order terms},\end{aligned}$$

and

$$r_1(t,t_a)=r_1(t,t-T)+[\partial r_1(t,t_a)/\partial t_a]_{t_a=t-T}(t_a-t+T)$$
$$+\text{higher order terms}.$$

The last term is of second order, since it is a product of two first-order quantities. Neglecting all second and higher order terms, one obtains

$$T-(t-t_a)=r_1(t,t-T)/r_0'(T). \quad (38)$$

For the corresponding dimensionless quantities, one obtains

$$\lambda-(\theta-\theta_a)=\lambda-\tau=S_1(\theta,\theta-\lambda)/S_0'(\lambda). \quad (39)$$

To obtain $A(R,t)$ substitute the value of t_a from Eq. (38) into $r''(t,t_a)$ and again use Taylor's expansion:

$$\begin{aligned}A(R,t)&=A(r_0(T),t)=[r''(t,t_a)]_{t_a=t-T+r_1(t,t-T)/r_0'(T)}\\&=[r_0''(t-t_a)+r_1''(t,t_a)]_{t_a=t-T+r_1(t,t-T)/r_0'(T)}\\&=[r_0''(T)-r_0'''(T)(r_1(t,t-T)/r_0'(T))+r_1''(t,t-T)]\\&\quad+\text{higher order quantities}. \quad (40)\end{aligned}$$

Neglecting higher order quantities, Eq. (35) reduces to

$$\begin{aligned}V(t)&=(m/e)\int_{R=r_c}^{R=r_p}A(R,t)dR\\&=(m/e)\int_{R=r_c}^{R=r_p}A(r_0(T),t)dr_0(T)\\&=(m/e)\int_{T=0}^{T=T_p}[r_0''(T)-r_0'''(T)(r_1(t,t-T)/r_0'(T))\\&\quad+r_1''(t,t-T)]r_0'(T)dT=(m/e)\int_{T=0}^{T=T_p}r_0''(T)r_0'(T)dT\\&\quad+(m/e)\int_{T=0}^{T=T_p}[r_0'(T)r_1''(t,t-T)\\&\quad-r_0'''(T)r_1(t,t-T)]dT=\frac{m}{e}\frac{r_c^2}{\alpha^2}\frac{(S_0')_p^2}{2}\\&\quad+\frac{m}{e}\frac{r_c^2}{\alpha^2}\int_{\lambda=0}^{\lambda=\lambda_p}[S_0'(\lambda)S_1''(\theta,\theta-\lambda)\\&\qquad-S_0'''(\lambda)S_1(\theta,\theta-\lambda)]d\lambda. \quad (41)\end{aligned}$$

The first term is the direct voltage; it is the same as that in Eq. (22) when no alternating current is superposed. The second term gives the variational component of the applied voltage. Dividing that by the superposed alternating current $j_1\epsilon^{i\omega t}=J_0K\epsilon^{p\theta}$, one obtains

$$z=\frac{1}{J_0K\epsilon^{p\theta}}\frac{m}{e}\frac{r_c^2}{\alpha^2}\int_0^{\lambda_p}[S_0'(\lambda)S_1''(\theta,\theta-\lambda)$$
$$-S_0'''(\lambda)S_1(\theta,\theta-\lambda)]d\lambda. \quad (42)$$

By Eq. (27), it follows that

$$S_1''(\theta,\theta_a)=\frac{K(\epsilon^{p\theta}-\epsilon^{p\theta_a})}{pS_0(\theta-\theta_a)}-\frac{S_0''(\theta-\theta_a)}{S_0(\theta-\theta_a)}S_1(\theta,\theta_a);$$

then

$$S_1''(\theta,\theta-\lambda)=\frac{K(\epsilon^{p\theta}-\epsilon^{p(\theta-\lambda)})}{pS_0(\lambda)}-\frac{S_0''(\lambda)}{S_0(\lambda)}S_1(\theta,\theta-\lambda).$$

Substituting this into Eq. (42), and recalling that $(S_0S_0'''+S_0'S_0'')=(S_0S_0'')'=1$, one obtains

$$z=\frac{1}{J_0K\epsilon^{p\theta}}\frac{m}{e}\frac{r_c^2}{\alpha^2}\int_0^{\lambda_p}\left[\frac{K}{p}\frac{S_0'(\lambda)}{S_0(\lambda)}(\epsilon^{p\theta}-\epsilon^{p(\theta-\lambda)})\right.$$
$$\left.-\frac{S_1(\theta,\theta-\lambda)}{S_0(\lambda)}\right]d\lambda. \quad (43)$$

When the expression for S_1 from Eq. (28) is substituted into the last equation, and the relation $S_0'''S_0+S_0'S_0''$

$=1$, is again used the impedance becomes

$$z=(-mr_c^2/J_0pe\alpha^2)\int_0^{\lambda_p}[\phi(\lambda)/S_0(\lambda)]d\lambda$$

$$=\frac{-mr_c^2}{J_0e\alpha^2}\int_0^{\lambda_p}\frac{C_0(\lambda)+C_1(\lambda)p+C_2(\lambda)p^2+\cdots}{S_0(\lambda)}d\lambda. \quad (44)$$

By Eq. (34), $C_0(\lambda)=-\frac{2}{3}S_0'(\lambda)\cdot\lambda$, and the first term in Eq. (44) is then

$$\frac{mr_c^2}{J_0e\alpha^2}\int_0^{\lambda_p}\frac{2}{3}\frac{S_0'(\lambda)\cdot\lambda}{S_0(\lambda)}d\lambda$$

$$=\frac{2mr_c^2}{3J_0e\alpha^2}\int_0^{\lambda_p}\frac{S_0'(\lambda)\cdot S_0(\lambda)S_0''(\lambda)}{S_0(\lambda)}d\lambda$$

$$=\frac{mr_c^2}{3J_0e\alpha^2}[S_0'(\lambda_p)]^2=\frac{mr_c^2}{3J_0e\alpha^2}[S_0']_p^2.$$

But this is $\frac{2}{3}$ of the dc resistance by Eq. (22a) and is equal to the variational plate resistance $(1/g)$ by Eq. (24). Thus, the leading term of the impedance is $1/g$. Equation (44) then becomes

$$Z=\frac{1}{g}\Bigg[1+\frac{1}{\int_0^{\lambda_p}C_0d\lambda/S_0}\times\int_0^{\lambda_p}\frac{pC_1(\lambda)+p^2C_2(\lambda)+p^3C_3(\lambda)+\cdots}{S_0(\lambda)}d\lambda\Bigg]$$

$$=\frac{1}{g}\Bigg[1-\frac{3}{(S_0')_p^2}\int_0^{\lambda_p}\frac{pC_1(\lambda)+p^2C_2(\lambda)+\cdots}{S_0(\lambda)}d\lambda\Bigg]. \quad (45)$$

Let k_1, k_2, k_3, k_4, etc., be defined as follows:

$$\left.\begin{aligned}
k_1(\lambda_p)&=\frac{10}{(S_0')_p^2\cdot\lambda_p}\int_0^{\lambda_p}\frac{C_1(\lambda)}{S_0(\lambda)}d\lambda,\\
k_2(\lambda_p)&=\frac{-45}{(S_0')_p^2\lambda_p^2}\int_0^{\lambda_p}\frac{C_2(\lambda)}{S_0(\lambda)}d\lambda,\\
k_3(\lambda_p)&=\frac{252}{(S_0')_p^2\lambda_p^3}\int_0^{\lambda_p}\frac{C_3(\lambda)}{S_0(\lambda)}d\lambda,\\
k_4(\lambda_p)&=\frac{-1680}{(S_0')_p^2\lambda_p^4}\int_0^{\lambda_p}\frac{C_4(\lambda)}{S_0(\lambda)}d\lambda,\\
k_5(\lambda_p)&=\frac{12{,}960}{(S_0')_p^2\lambda_p^4}\int_0^{\lambda_p}\frac{C_5(\lambda)}{S_0(\lambda)}d\lambda, \text{ etc.}
\end{aligned}\right\} \quad (46)$$

Equation (45) becomes

$$Z=\frac{1}{g}\Big[1-i\frac{3}{10}\beta k_1-\frac{1}{15}\beta^2k_2+\frac{i}{84}\beta^3k_3+\frac{\beta^4}{560}k_4-\frac{i}{4320}\beta^5k_5-\frac{\beta^6}{37{,}800}k_6+\cdots\Big], \quad (47)$$

where β is the transit time angle defined as

$$\beta=\omega Tp=\omega\alpha\lambda_p. \quad (48)$$

The integrals

$$\int_0^{\lambda_p}(C_1/S_0)d\lambda, \quad \int_0^{\lambda_p}(C_2/S_0)d\lambda, \text{ etc.},$$

are evaluated by numerical integration; and from these integrals, the coefficients k's are calculated, using Eq. (46). The results for the first five coefficients are plotted against $S_0=r_p/r_c$ in Fig. 5. For r_p/r_c nearly unity, all the k's approach unity, and Eq. (47) reduces to the result derived by Llewellyn for the parallel plane case.[3]

In all computations leading to the final results of k's enough significant figures are carried for an accuracy of ±0.001. However, there are inevitably accidental slight errors. It is estimated that all k's are accurate to ±0.003. The maximum error due to dropping off of terms beyond the sixth in Eq. (47) can be estimated by the magnitude of the seventh term $(\beta^6k_6/37{,}800)$. It can be proved with some complicated mathematical manipulation that the coefficients $k_n(\lambda_p)$ decrease both with n and λ_p. For $\beta\leqslant\pi$, $\beta^6/37{,}800<0.026$, therefore, the first term dropped off is only 0.026 of the leading term, for r_p nearly equal to r_c, and is much smaller for larger r_p/r_c. The actual error is even less because of the alternations in signs in both the resistive and the reactive terms.

The following is a numerical example of the application of the results derived in this work: Let

r_c = cathode radius = 0.5 mm = 0.000,5m,
r_p = plate radius = 3.0 mm = 0.003m, $r_p/r_c=6.00$,
1 = length of tube = 2 cm = 0.02m
J_0 = 5 ma/cm = 0.5 amp/m.

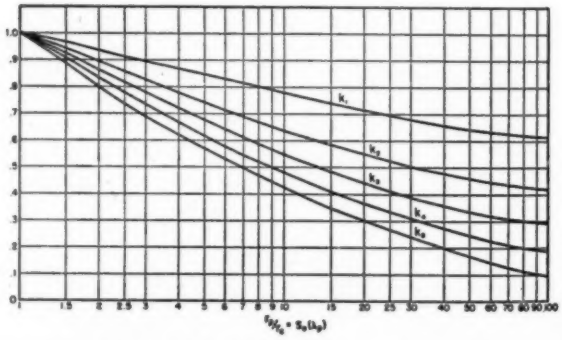


FIG. 5. Impedance coefficient, k vs r_p/r_c.

From Fig. 1, $\tau=3.76$ for $S_0(\tau)=6.00$. Therefore, one has $\lambda_p=3.76$; and from Fig. 3, $(S_0')_p=2.80$. By Eq. (13), one obtains

$$\alpha=6.81\times10^{-8}[r_c^2/J_0]^{\frac{1}{2}}=5.40\times10^{-10}\text{ sec},$$
$$T_p=\alpha\lambda_p=3.76\times5.40\times10^{-10}=2.03\times10^{-9}\text{ sec}.$$

The steady voltage required is by Eq. (22)

$$V_0=\tfrac{1}{2}(m/e)(r_c^2/\alpha^2)(S_0')_p^2=19.1\text{ volts}.$$

The dc resistance is then

$$R_0=V_0/J_0=19.1/0.5=38.2\Omega/\text{meter}.$$

The variational plate resistance is

$$1/g=\tfrac{2}{3}R_0=25.5\Omega/\text{meter}.$$

From Fig. 5, for $S_0(\lambda p)=6$, one has

$$k_1=0.833,$$
$$k_2=0.716,$$
$$k_3=0.642,$$
$$k_4=0.582,$$
$$k_5=0.538.$$

At a frequency of 200 mc, one has

$$\beta=\omega Tp=2\pi\times2\times10^8\times2.03\times10^{-9}$$
$$=2.55\text{ radians}.$$

Substituting these values into Eq. (47), one obtains

$$Z=25.5\left[\left(1-\frac{1}{15}\beta^2k_2+\frac{1}{560}\beta^4k_4\right)-i\left(\frac{3}{10}\beta k_1-\frac{1}{84}\beta^3k_3+\frac{1}{43{,}200}\beta^5k_5\right)\right]$$
$$=(18.7-i13.4)\Omega/\text{meter}.$$

For tube of length 2 cm, one obtains

$$\text{impedance}=Z/0.02=935-i670\Omega.$$

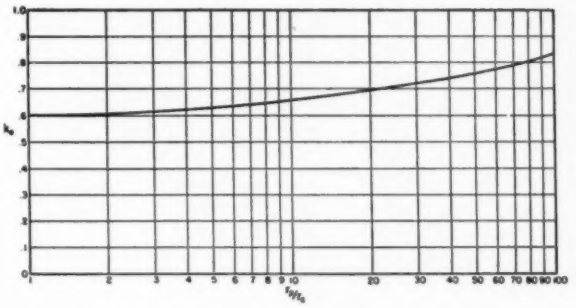


FIG. 6. Effective inductive capacity, k_e vs r_p/r_c.

VII. THE INTERELECTRODE CAPACITY AT VERY LOW FREQUENCY

At very low frequency the impedance can be regarded as that of the variational plate resistance shunted by interelectrode capacity. Retaining only the first two terms of Eq. (47) for very small β, and taking the reciprocal of the impedance, and neglecting higher powers of β, one obtains

$$Y=1/Z=g[1+i(3/10)\beta k_1]=g[1+i(3/10)\omega\alpha\lambda_pk_1].$$

The equivalent capacitance is then

$$C_e=3\alpha\lambda_pgk_1/10\text{ farad/meter}. \tag{49}$$

The equivalent specific inductive capacity of the interelectrode space is the ratio of this capacity to that between the same two electrodes in vacuum.

$$k_e=C_e/[2\pi\epsilon_0/\ln(r_p/r_c)]$$
$$=(9/10)[\lambda_pk_1/(S_0')_p^2]\log(S_0)_p \tag{50}$$

The value of k_e has been calculated and is shown in Fig. 6. It starts at 0.6 for very low r_p/r_c, agreeing with Llewellyn's result for the parallel plane diodes. It increases toward unity for very large r_p/r_c. This is understandable, because, for large r_p/r_c, the region contributing most to the impedance has very low electron density.

Free Longitudinal Vibrations of Rubber and Tissue Strips*

RICHARD W. LAWTON AND ALLEN L. KING
Cornell University Medical College, New York, New York, and Dartmouth College, Hanover, New Hampshire
(Received May 11, 1951)

Frequencies of free longitudinal vibrations of several different specimens of rubber and aortic strips were measured as a function of applied load and relative length L/L_0, where L_0 is the initial length. Curves of $\omega^2 L_0/g$ *versus* L/L_0 have distinct minima that become sharper and move upward and toward the $\omega^2 L_0/g$ axis as the specimens become stiffer. The experimental curves are described rather well by theoretical equations derived from current relations for the extension of an ideal elastomer; they may be used for determining values of the elastomeric parameter β. The method is simple and rapid and is applicable to biological tissues as well as to other rubberlike substances.

THE bulk of available information on the dynamic behavior of rubber and rubberlike substances has been obtained from loaded specimens under forced vibration and over a wide range of frequencies. The results have been treated in terms of conventional spring and dashpot models or in terms of electrical analogs. Rather few measurements have been made of free longitudinal vibrations, perhaps in part because of the lack of an adequate theoretical treatment of the problem. A few preliminary free-vibration experiments on rubber specimens suggested that valuable data may be obtained in this manner for determining elastomeric parameters. We shall describe here a study of the free longitudinal vibrations of several types of rubber and of tissue strips from young and old aortas.

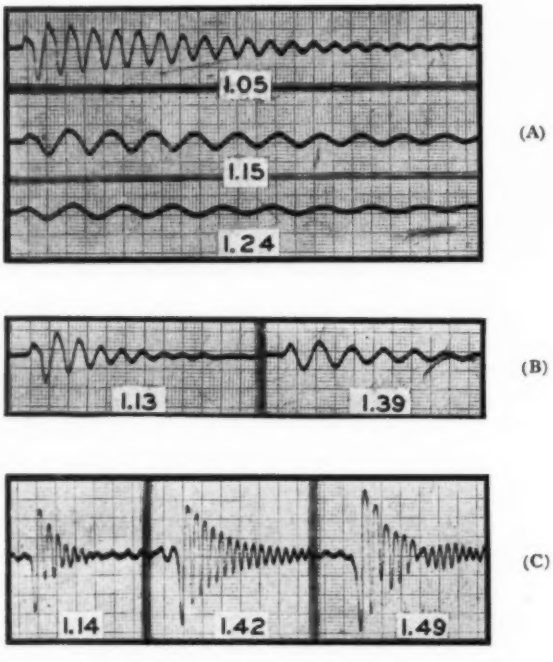


FIG. 1. Galvanometer tracings recorded from piezoelectric crystal. (A) Gum rubber, $L_0=27$ cm, (B) filled gum stock (50 percent medium processing channel black), $L_0=21.1$ cm, and (C) a circumferential strip from thoracic aorta of a dog, $L_0=1.8$ cm. Values of α given below the records. Least count of time scales 0.04 sec. Note beating in the third record of (C).

The rubber strips for many of these experiments were cut from thin sheets by means of an I-shaped die having a center bar one inch long.† Their thickness varied from 0.005 to 0.01 inch. A few rubber strips were cut to a length of more than 20 cm. Light spring clamps were attached to each end of a strip and the distance between the clamps was measured by means of calipers. This distance was taken to be the initial length L_0. The strip then was suspended from a phonograph pickup, and weights were hooked to the lower clamp. Relative changes in length were found by measuring the distance between two marks near the center of the strip for each applied load and by dividing it by the distance between these marks for the unstretched state.

The output of the pickup was amplified and recorded on a direct-writing (heated stylus) galvanometer of the electrocardiographic type.‡ This instrument has an upper frequency limit of 80 cycles per second, and it appears to be entirely satisfactory for obtaining frequency records of low frequency events. Ballou and Smith[1] have obtained similar records for textile fibers with a photoelectric device. Some frequencies were so low, especially for the long specimens, that they were measured by counting vibrations and timing them with a stopwatch. Two methods were found satisfactory for starting the vibrations. In the more satisfactory one, the weighted specimen was stretched by pulling on a thread attached to the weight; and, when the recorder indicated a smooth baseline, it was released quickly. In the other, the weights were lifted momentarily and released. Measurements were made of the frequencies of free longitudinal vibrations for a specimen by increasing the load in a continuous series of steps without removing the load between observations. Relative extended lengths were obtained from measurements of the distance between the two marks by means of a precision cathetometer. All experiments reported here were performed at room temperature.

* This investigation was supported in part by a research grant from the National Institutes of Health, Public Health Service.

† The rubber specimens were kindly supplied by the General Laboratories, U. S. Rubber Company, Passaic, New Jersey.

‡ Made by the Sanborn Company, Cambridge, Massachusetts.

[1] J. W. Ballou and J. C. Smith, J. Appl. Phys. **20**, 493 (1949).

Typical galvanometer records are shown in Fig. 1, where the least count of the time scales is 0.04 second. Sometimes an apparent beating effect occurred, especially in the biological materials, as shown in Fig. 1(C). Beating often was observed for a whole rubber band, probably because of inequalities in the length, composition and other factors for its two sides. The same explanation may very well apply to the beating in the biological materials.

EXPERIMENTAL RESULTS

If there is no dissipation and if the amplitude is small, a mass suspended from a strip of rubber or rubberlike substance may oscillate with simple harmonic motion at a frequency $\omega/2\pi$ given by the relation

$$\omega^2 L_0/g=(\alpha^3+2)/\alpha(\alpha^3-1), \qquad (1)$$

where g is the gravitational acceleration, L_0 is the unstretched length of the specimen, and α equals the ratio of the extended to the unstretched length L/L_0. This relation is derived from Wall's load-length equation for an elastomeric strip.[2] As seen in Fig. 1 the motion actually is damped with a logarithmic decrement of the order 0.05. The frequency of the damped oscillation can be shown to be only about 0.02 percent less than the undamped frequency. For this reason the damping effect is neglected and ω is assumed to correspond to the measured frequency. Note that $\omega^2 L_0/g$ is a universal function of α.

In Fig. 2 are shown experimental results on two gum rubber stocks A and B, a strip cut from a household rubber band C, a filled gum stock D, and a circumferential strip of human aorta E obtained post mortem. The foregoing universal function also is plotted for comparison. All the experimental curves lie above the universal one for the most part and approach it as α decreases towards unity. Even the gum rubber stocks show definite minima, although they agree rather well out to nearly 200 percent elongation. The filled sample D contains 50 percent by weight of medium processing channel black per hundred of gum stock A, and it is definitely stiffer than the untreated gum stock. The minimum in the frequency response curve becomes sharper and rises to the left for substances of increasing stiffness. The aortic strip was cut as a ring from the thoracic region of a 21-year-old male post mortem and opened to form the strip. Although beating was present in some of the aortic frequency records, only the mean frequencies are plotted. The aortic strip appears to be stiffer than any of the rubber specimens.

In Fig. 3 the foregoing aortic strip from a 21-year-old male is compared with a similar strip from a 70-year-old-male. As may be expected, the old aorta is stiffer than the young one. In a previous paper[3] the suggestion was made that aging arterial walls may lose their extensibility as a result of increased interchain bonding and deposition of filler particles.

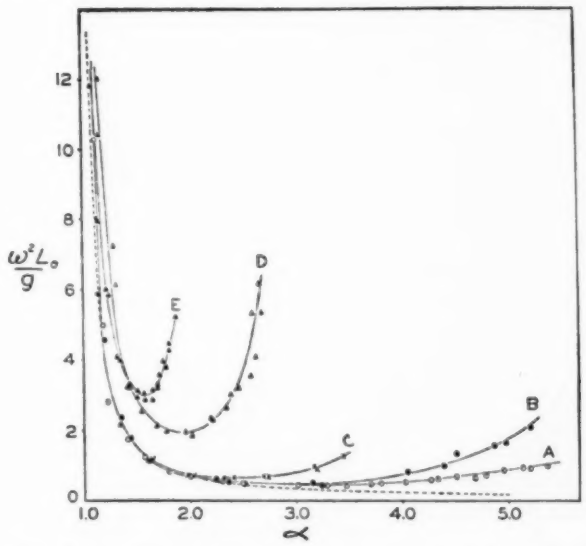


FIG. 2. Experimental frequency response curves. (A) Gum stock (medium sulfur content), $L_0=2.5$ cm, (B) gum stock (high sulfur content), $L_0=2.6$ cm, (C) strip from household rubber band, $L_0=2.8$ cm, (D) gum stock A with an active filler (50 percent medium processing channel black), $L_0=2.6$ cm, (E) circumferential strip of human thoracic aorta (21 years), $L_0=2.86$ cm. Universal frequency response curve plotted as a dotted line.

All the rubber specimens of Fig. 2 were given a preliminary treatment by being stretched to 400 percent of their initial lengths. Such treatment does not alter the frequency response curve markedly, as can be seen in Fig. 4. Here are shown the results on a long specimen of gum stock rubber which was carried through the series of measurements three times, then given the stretching treatment, and finally carried through the series once more. All the data agree fairly closely, although there is a slight indication that the treatment

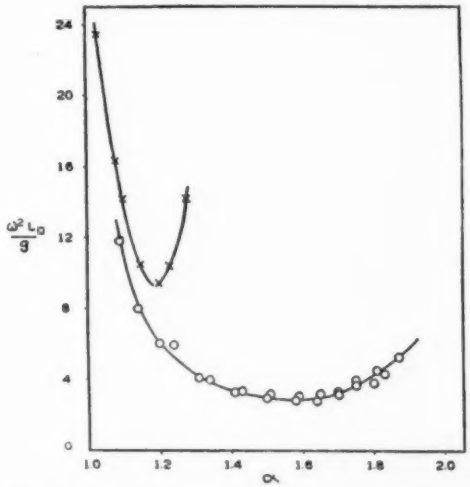


FIG. 3. Influence of age on frequency response of aortic strips. Open circles: 21-year-old male, $L_0=2.86$ cm; crosses: 70-year-old male, $L_0=4.6$ cm.

[2] F. T. Wall, J. Chem. Phys. **10**, 485 (1942); **11**, 527 (1943).
[3] A. L. King, J. Appl. Phys. **17**, 501 (1946).

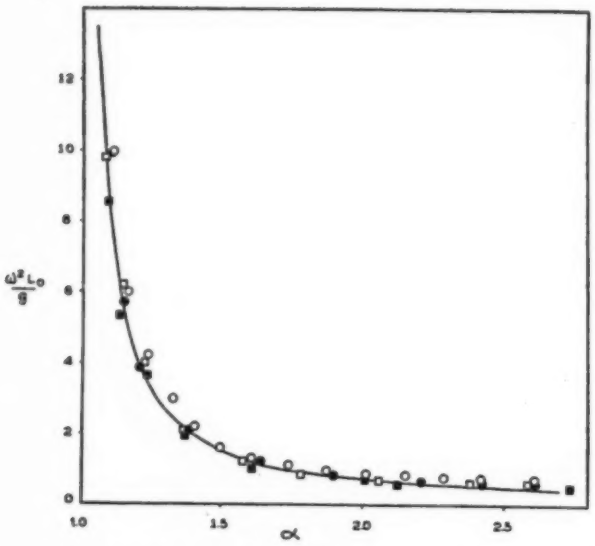


FIG. 4. Multiple extensions of single gum stock specimen. At start $L_0=22.4$ cm. Open circles, open squares, and solid circles are first, second, and third extensions, respectively; solid squares are for data after ≈ 400 percent extension. The solid line is the simple universal frequency response curve.

brings about somewhat better agreement with the universal curve.

DISCUSSION

The divergence between the experimental data and the simple universal curve with increasing extension and especially the appearance of a minima needs explanation. As a matter of fact, it is this very divergence which allows one to determine a fundamental elastomeric parameter for the specimen material. If the dissipation may be neglected and if the amplitudes are not too large, the James-Guth model of a rubber strip[4,5] can be shown to yield the function

$$\omega^2 L_0/g=(D\beta\alpha^3+\tfrac{1}{2}B\beta+\tfrac{3}{2}u\alpha^{\frac{1}{2}})/(v\alpha^3-u\alpha^{\frac{1}{2}}), \qquad (2)$$

where α is the relative length L/L_0 and β is the elastomeric parameter. This dimensionless parameter may be interpreted to equal the ratio of the distance between the ends of a chain in the unstretched state to the distance between these same ends when the chain is fully extended. The other symbols have the following meanings:

$$1/B=(1/u^2)-1/\sinh^2 u, \qquad 1/D=(1/v^2)-1/\sinh^2 v,$$

and u and v are the inverse Langevin functions of the arguments $\beta\alpha^{-\frac{1}{2}}$ and $\beta\alpha$, respectively. Equation (2) reduces to Eq. (1) for β equal to zero. Values of $\omega^2 L_0/g$ as determined from Eq. (2) are plotted against α in Fig. 5 for several values of β. This family of curves looks very much like the series of experimental curves shown in Fig. 2.

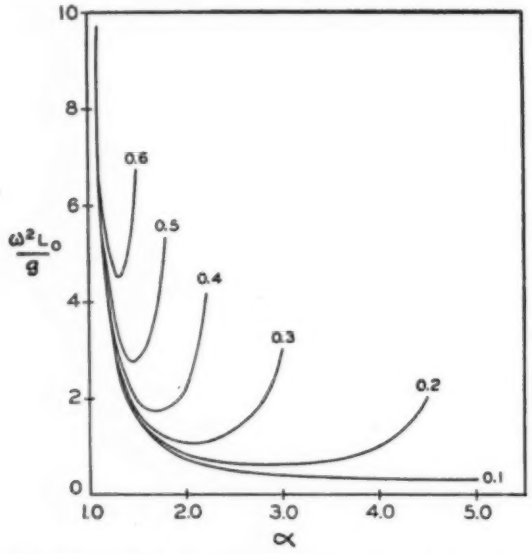


FIG. 5. Theoretical frequency response curve of an ideal elastomer for values of the parameter β from 0.10 to 0.60. Curve for $\beta=0.0$ not shown. It approximates $\beta=0.10$ over the range of values of α shown.

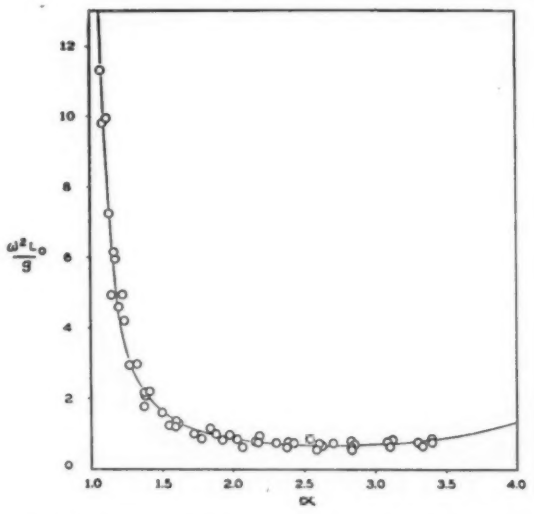


FIG. 6. Frequency response of several specimens of gum stock (A). L_0 averaged 22.0 cm. The solid line represents Eq. (2) with $\beta=0.21$. Note that both theory and experiment have the same minimum value for $\omega^2 L_0/g$ and α.

The agreement between theory and experiment for several specimens of a typical gum stock is shown in Fig. 6. Here β equals 0.21 and the value of $\alpha_{\min}$ for both the theory and the experiment coincide at 2.60. For stiffer rubbers and the aortic strips, however, the agreement in detail appears poor. The nature of the disagreement is shown in Fig. 7 for the rubber strip containing channel black as an active filler. The theoretical curve with β equal to 0.40 has a minimum value of $\omega^2 L_0/g$ near that observed for the strip, but it appears to be displaced to one side. The reason for this displacement is not entirely clear. Perhaps it arises in part because the complex filled specimen is not described adequately

[4] H. M. James and E. Guth, J. Chem. Phys. 11, 455 (1943).
[5] A. L. King, Am. J. Phys. 14, 28 (1946).

by the simple theory for an ideal elastomer. Swelling appears to reduce this discrepancy markedly; the observations on swelling are to be published in another paper. Preliminary experiments indicate that the agreement with theory also is enhanced if data are obtained during progressive unloading of the unswollen specimen from an extended length. In this case L_0 is the final unstretched length.

We believe, however, that the characteristic elastomeric parameter β may be determined for any unswollen specimen by selecting the theoretical curve with a minimum value for $\omega^2 L_0/g$ equal to that of the experimental data. If this procedure is permissable, the aortic strip from the young male is found to have a β of approximately 0.50 and that from the old male a β of 0.76. These dynamic values of β are significantly higher than those obtained by static methods.[3] Such differences are commonly found for elastomers. Furthermore, whenever an elastomeric specimen is cut, chains are shortened and β increases in value. On the other hand, the observed differences may be entirely the result of random variations.

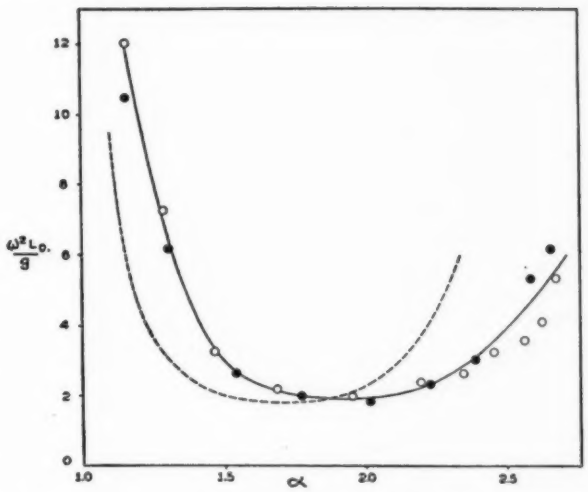


FIG. 7. Frequency response curve for two specimens of the same filled gum stock. The dotted line is the theoretical curve with $\beta=0.40$.

APPENDIX I

The realtion between applied force to stretch a strip of rubber and the stretched length was first developed by Wall[2] from statistical considerations. It is

$$F=k(\alpha-1/\alpha^2), \qquad (1)'$$

where α equals the ratio L/L_0 and k is a physical constant which depends upon the absolute temperature and the elastomeric properties of the specimen. If a weight W is hung from a rubber strip, therefore, it will produce a relative stretched length α. Any additional small force ΔF increases the relative length by $\Delta L/L_0$, where for small elongations

$$\Delta F=k(1+2/\alpha^3)\Delta L/L_0. \qquad (2)'$$

The restoring force exerted by the rubber strip thus is proportional to the elongation. The system can oscillate with simple harmonic motion of frequency $\omega/2\pi$, where

$$\omega^2=k(1+2/\alpha^3)g/WL_0. \qquad (3)'$$

But from Eq. (1) k/W equals $1/(\alpha-1/\alpha^2)$ so that on substituting and rearranging, we have

$$\omega^2 L_0/g=(\alpha^3+2)/\alpha(\alpha^3-1). \qquad (4)'$$

This is the equation for the simple universal frequency response curve.

APPENDIX II

The force exerted by an elastomeric strip has been shown to have the form[5]

$$F=A[\mathfrak{L}^{-1}(\beta\alpha)-\alpha^{-1}\mathfrak{L}^{-1}(\beta\alpha^{-\frac{1}{2}})] \qquad (1)''$$

over a wide range of values for the relative length α. Here β is the elastomeric parameter and A is a constant for a substance under isothermal conditions. The functions $\mathfrak{L}^{-1}(\beta\alpha)$ and $\mathfrak{L}^{-1}(\beta\alpha^{-\frac{1}{2}})$ are inverse Langevin functions. The Langevin function of u is represented by $\mathfrak{L}(u)$ and is equal to $\coth u-1/u$. Its inverse is u.

An expression for a small additional force ΔF may be found from Eq. (1)'' by differentiating with respect to α. The result is

$$\Delta F=A[\beta D+\tfrac{3}{2}u\alpha^{-5/2}+\tfrac{1}{2}\beta B\alpha^{-3}](\Delta L/L_0), \qquad (2)''$$

where B and D have the meanings given in Eq. (3) of the text and u equals $\mathfrak{L}^{-1}(\beta\alpha^{-\frac{1}{2}})$. The system can oscillate with simple harmonic motion of frequency $\omega/2\pi$, where

$$\omega^2=A[\beta D+\tfrac{3}{2}u\alpha^{-5/2}+\tfrac{1}{2}\beta B\alpha^{-3}]g/WL_0. \qquad (3)''$$

Finally F in Eq. (1)'' is replaced by W and the value of A/W from the equation is substituted in Eq. (3)''. The result is

$$\omega^2\mathfrak{L}_0/g=(D\beta\alpha^3+\tfrac{1}{2}B\beta+\tfrac{3}{2}u\alpha^{\frac{1}{2}})/(v\alpha^3-u\alpha^{\frac{1}{2}}), \qquad (4)''$$

where v represents $\mathfrak{L}^{-1}(\beta\alpha)$. This is Eq. (2) of the text.

The Application of Nutting's Equation to the Viscoelastic Behavior of Certain Polymeric Systems

ROLF BUCHDAHL AND LAWRENCE E. NIELSEN
Research Department, Plastics Division, Monsanto Chemical Company, Springfield, Massachusetts
(Received May 11, 1951)

The Nutting equation $\epsilon=\psi\times\sigma^{\beta}\times t^{n}$, where ϵ=strain, σ=stress, t=time, and ψ, β, and n are material constants, is applicable to a large number of polymeric systems, such as plasticized polyvinyl chloride, styrene-butadiene copolymers, polystyrene. It has been shown that the material constant n is closely related to the damping coefficient as obtained from dynamic mechanical measurements The temperature dependence of n and δ (damping coefficient) are about the same; and, in some cases, there exists a simple quantitative relationship between the two constants. A mathematical derivation of the relationship between n and δ is given and a physical interpretation of the dimensional changes in the Nutting equation, in terms of complex moduli, is presented. Furthermore, it is shown that, in the case of styrene-butadiene copolymer, n varies with the differences in gel content and gel structure.

INTRODUCTION

THE phenomena of creep, stress relaxation, and recovery are observable in metals, crystals, glasses (organic or inorganic), rubbers, etc. The viscoelastic behavior of solid organic polymers only differs from that of other material in that large variations in these properties can be brought about by relatively small changes in temperature and/or composition. A theoretical interpretation of these phenomena, in terms of molecular processes, requires, first of all, a satisfactory mathematical representation of the large number of available experimental data. Two methods have been suggested to obtain useful parameters to characterize the deformation processes. One might be termed the analytical approach:[1] The over-all deformation process is separated into two or three processes, as shown in Eq. (1). The integral approach does not attempt such a separation and is based on a relationship first suggested by Nutting,[2] as shown in Eq. (2).

$$\epsilon/\sigma=1/E_0+1/E_i(1-\exp(-t/r_i))+(1/\eta)t. \quad (1)^*$$

$$\epsilon=\psi\sigma^{\beta}t^{n}. \quad (2)^*$$

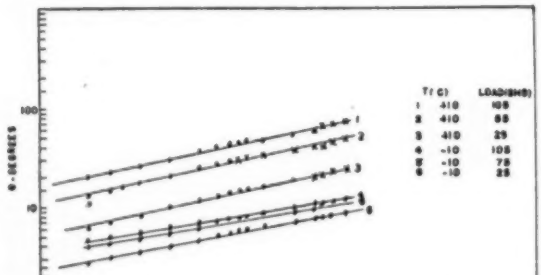


FIG. 1. Deformation-time curves for a polyvinyl chloride-plasticizer composition at two different temperatures and three different shear stresses.

During the last few years the analytical method has been used extensively in the characterization of the viscoelastic behavior of polymeric systems. As more and more experimental data became available, it was necessary to abandon the inherent simplicity of this method by the introduction of a relaxation spectrum. This necessity not only limits the usefulness of this method to characterize the viscoelastic behavior by a small number of intrinsic parameters, but it has also raised new questions relating to the physical interpretation of such a spectrum. The integral method, based on the Nutting equation, has been given relatively scant attention in the polymer field, although it has been used extensively to characterize deformation processes of such varied materials as metals, asphalts, and baking dough. The general objections raised in opposition to the use of this method have been: (a) that the time dependence of the viscoelastic deformations cannot always be expressed by a single constant; (b) that the variation of the exponent "n" between 0 and 1 leads to varying dimensions in the material constant ψ; (c) that the constant "n" is an arbitrary quantity which cannot be related to other quantities of the material. In the following it will be shown that these

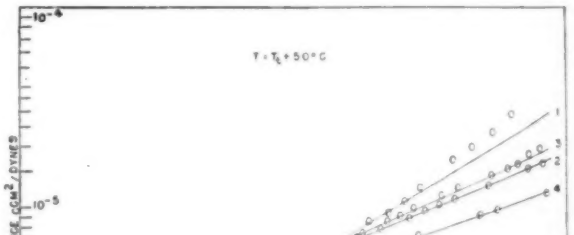


FIG. 2. Deformation-time curves for different styrene-butadiene copolymers. T_t=temperature of maximum damping.

[1] T. Alfrey, Jr., *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York, 1948).
[2] P. Nutting, Proc. Am. Soc. Testing Materials **XXI**, 1162 (1921).
* ϵ=strain, σ=stress, t=time, E_0 and E_i=moduli, η=viscosity, ψ, β, and n=constants.

objections are not particularly valid and that the Nutting equation not only gives a simple characterization of the viscoelastic behavior of polymers but promises to result in a better understanding of the molecular mechanism responsible for this type of deformation process.

THE TIME DEPENDENCE OF VISCOELASTIC DEFORMATIONS

According to the Nutting equation the strain, at constant stress, is proportional to t^n where n is assumed to be a constant at a given temperature. Figures 1, 2, and 3 show some results obtained on various polymer systems and it can be seen that the relationship is of the form given by the Nutting equation. (Many more creep curves have been analyzed by this procedure and found to give a linear relationship between log strain and log time; see also the papers by Hahn and Gardik,[3] and Scott, Blair, and Copper.[4]) It should also be pointed out that the data in Figs. 1, 2, and 3 were obtained on different instruments.

Other investigators[5,6] have reported data which give a linear relationship between the strain and the logarithm of time. For small strains such a relationship follows from the Nutting equation by expanding the logarithm of strain into a power series. There are, however, experimental data which cannot be represented by a single constant n, as is shown in Fig. 4. Usually such data can be approximated quite well by two curves of constant "n" of which one curve will have a slope equal or rather close to one. Systems exhibiting this kind of viscoelastic behavior possess a semipermanent network.[7] The quantity "n" which characterizes the nature of the time dependent deformation process can only remain constant if the molecular processes remain unchanged during the time of the test. If they change, "n" must change by necessity. The fact that "n" does not always remain constant during a long-time creep test should, therefore, not be used as an argument against the applicability of the Nutting equation to systems in which the network points, or crosslinks, are not of permanent kind.

There can be no doubt, therefore, that the Nutting equation can be applied to characterize quantitatively the creep behavior of a large number of polymeric compositions. The determination of the parameter characteristics of the material—particularly n, but also ψ and β—is simple and more readily accomplished than by the use of the analytical approach. One other point is worth emphasis: Using the analytical approach one frequently calculates a viscosity term from the linear portion of the deformation-time curve. Quite often such a calculation is carried out on curves having an "n"

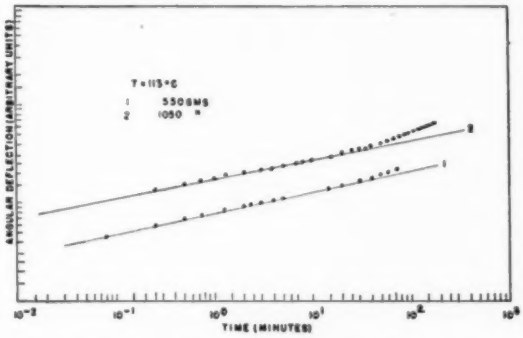


FIG. 3. Deformation time curves for polystyrene at two different shear stresses.

considerably smaller than one. As a viscosity, by definition, can only be determined when $n=1$ it is obvious that such data are completely misleading. In general one finds that temperature coefficients of viscosities obtained from such data are unusually high[8] which, in turn, has led to a considerable amount of speculation concerning the origin of such a high "apparent" activation energy of viscous flow of polymers. This difficulty disappears when one accepts as viscosity values only such data which, on a $\log\epsilon - \log t$ plot, have a slope of not less than about 0.9.

THE TEMPERATURE DEPENDENCE OF "n" AND "ψ"

It was pointed out some time ago[9] that the shape of the creep curve of plasticized polyvinyl chloride compositions varied in a systematic manner with the temperature; similar conclusions could also be drawn from creep experiments made by Conant and Liska[10] on various rubbers. These observations suggested a more careful investigation of the temperature dependence of parameters of the Nutting equation. For this reason, the creep of two plasticized polyvinyl chloride compositions was investigated in detail over a wide temperature range. Some of the experimental data are shown in Fig. 1 and the values of n and ψ at all tem-

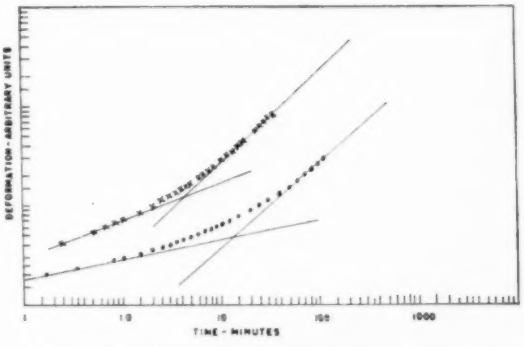


FIG. 4. Deformation-time curves for polystyrene.

[3] S. H. Hahn and I. Gardik, Rubber Chem. and Tech. **14**, 433 (1941).
[4] Scott, Blair, and Copper, J. Soc. Chem. Ind. **60**, 190 (1941).
[5] W. Brenschede, Kolloid Z. **104**, 1 (1943).
[6] Conant, Hall, and Lyons, J. Appl. Phys. **21**, 499 (1950).
[7] L. E. Nielsen and R. Buchdahl, J. Colloid Sci. **5**, 282 (1950).
[8] G. J. Dienes, Meeting of Soc. of Rheology, New York, New York, November, 1950.
[9] Nielsen, Buchdahl, and Levreault, J. Appl. Phys. **21**, 607 (1950).
[10] F. S. Conant and J. W. Liska, J. Appl. Phys. **15**, 767 (1944).

TABLE I. n, ψ, and E''/E' as a function of temperature.

Temp. (°C)	PVC+0.417ϕ, DOP n	ψ[a]	E''/E'	PVC+0.378ϕ, TCP n	ψ	E''/E'
50	0.080	2.5	0.095	0.085	3.6	0.127
40	0.075	1.4	0.133	0.087	3.0	0.143
30	0.115	1.1	0.209	0.133	1.5	0.318
20	0.158	0.64	0.30	0.355	0.25	0.66
10	0.220	0.43	0.41	0.25	0.066	0.45
0	0.266	0.17	0.44	0.11	0.0082	0.175
−10	0.175	...	0.24	0.047	0.016	...
−20	0.113	0.12	0.155	0.027	0.010	...
−30	0.090	0.03	0.150	0.026	0.006	...
−40	0.073	0.02	...	...	...	
−50	0.060	0.018	...	0.024	0.005	...

[a] ψ in arbitrary units.

peratures are given in Table I (see also Fig. 5). The results obtained are in complete agreement with data given by Bilmes[11] for polyvinyl chloride samples of unknown composition and one is justified, therefore, to consider the relationship between n and T and ψ and T applicable to polymer compositions of this kind irrespective of the particular plasticizer or polymer used. It was also found that "n" was independent of stress within the range of stresses used in these experiments. The accuracy of the ψ-value is considerably less than the accuracy of n because a double extrapolation is necessary to obtain ψ, whereas n is obtained directly from the $\log\epsilon - \log t$ curve. As β is rather close to unity for these materials the creep behavior over an extended time interval is completely characterized by the two constants n and ψ. The analytical method, on the other hand, requires a very much larger number of constants to describe adequately the viscoelastic behavior of these materials.

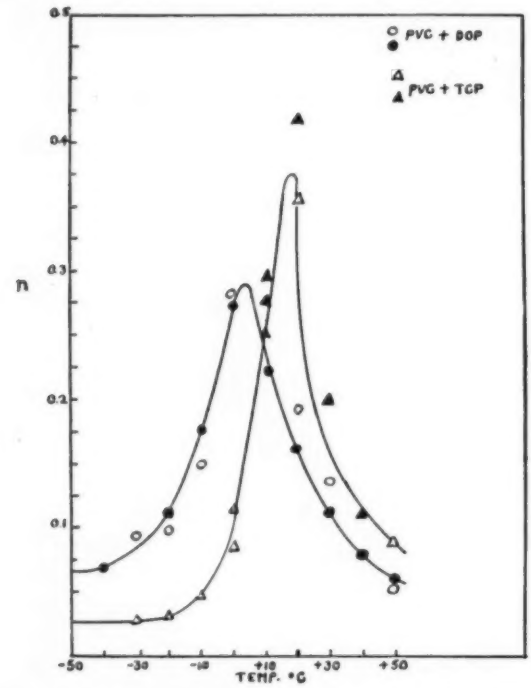


FIG. 5. n as a function of temperature for two plasticized polyvinyl chloride compositions (see Table I; ○, ▲ are calculated from the dissipation factors, see reference 12).

[11] L. Bilmes, J. Soc. Chem. Ind. 63, 182 (1944).

THE RELATIONSHIP OF "n" AND "ψ" TO OTHER PHYSICAL QUANTITIES

The shape of the $n-T$ curve and also the dependence of ψ on T are qualitatively very similar to the curves of the damping or dissipation factor and of the dynamic modulus as a function of temperature.[9] At the temperature of maximum damping the constant "n" goes through a maximum and on either side of the maximum "n" decreases to fairly small values. Furthermore, the width of the two curves ($n-T$ and $E''/E'-T$) are also related as can be seen from Fig. 5; a broad dissipation maximum yields a broad maximum for the $n-T$ curve. This striking qualitative similarity between the two independent physical parameters n and E''/E' suggests very strongly that a quantitative relationship should be obtainable without reference to specific molecular models. Indeed, a quantitative relationship can be readily obtained starting from an equation given by Zener[12] and combining it with the Nutting equation. Zener showed that the following equation relates the dissipation coefficient to the stress relaxation function:

$$E''/E' \doteq -\pi/2(d\ln\sigma/d\ln t), \tag{3}$$

where E' and E'' are the real and imaginary part of the complex modulus E^* ($E^*=E'+iE''$). This relationship only presupposes the validity of the Boltzmann superposition principle. If the further assumption is made that the stress relaxation function $\sigma(t)$ and the creep function $\epsilon(t)$ obey the relation

$$\epsilon(t)=1/\sigma(t), \tag{4}$$

then

$$\frac{d\ln\sigma}{d\ln t}=-\frac{d\ln\epsilon}{d\ln t}. \tag{5}$$

Equation (3) can then be written as

$$\frac{E''}{E'}\doteq-\frac{\pi}{2}\frac{d\ln\epsilon}{d\ln t}. \tag{6}$$

Putting Eq. (1) into logarithmic form and differentiating with respect to time gives

$$d\ln\epsilon/d\ln t=n. \tag{7}$$

Combining (7) and (6) gives the desired relationship between n and the dynamic dissipation factor

$$n\doteq 2E''/\pi E'. \tag{8}$$

How closely this relationship agrees with the experimental data can be seen from Fig. 5 where the dissipa-

[12] C. Zener, *Elasticity and Anelasticity of Metals* (The University of Chicago Press, Chicago), p. 54; also T. S. Kê, Phys. Rev. 71, 533 (1947).

tion factors—given in Table I—have been converted into n values. (The dynamic dissipation factors have been reported previously;[9] for the tricresyl phosphate-polyvinyl chloride compositions E''/E' has been corrected to take into account small differences in plasticizer concentration between the specimens used in creep and dissipation factor measurements.) Similarly, a close correlation can be obtained between the dynamic shear modulus (E^*) and the 5 or 10 second modulus calculated from a creep curve, as can be seen from Fig. 6.† This means then that the constant ψ of the Nutting equation is proportional to the reciprocal dynamic modulus for these and similar polymeric compositions. It is possible, therefore, to calculate with reasonable accuracy the complete creep curve from dynamic mechanical data or vice versa, provided the relationships given above are fulfilled.

It was pointed out above that the product of creep function and the stress relaxation function must be equal to one, in order for Eq. (8) to hold. It has been shown that this is at least approximately so for many polymeric compositions.[13,14] However, this condition does not hold in all cases: The creep of polystyrene above its second order transition temperature is an example of this (see also Fig. 4). Whenever the creep function contains a large amount of nonrecoverable deformation, Eq. (8) can only be valid for that part of the creep function which is identical with the stress-relaxation or recovery function.[14]

THE DIMENSIONAL ANALYSIS OF THE NUTTING EQUATION

One of the most serious objections to the Nutting equation is based on the fact that the use of the fractional and variable exponent n results in the constant ψ having variable dimensions, as the dimensions of ϵ, σ, and t are fixed. Although Scott Blair[15] has argued that the use of constants of variable dimensions is justified from various points of view, it would be most desirable to give a satisfactory explanation of this fact within the framework of classical physics which requires that the dimensions of physical parameters remain constant. Using the complex dynamic modulus E^* it is possible to understand the variations in dimensions of ψ.

Let us assume that the creep function $\epsilon(t)$ is given by the following equation (assuming $\beta=1$)

$$\epsilon=\frac{\sigma}{E^*}=\frac{\sigma}{E'+iE''}. \tag{9}$$

It can be shown that $E''=\omega\eta=2\pi\eta/t$. Inserting this expression into (9) gives

$$\epsilon=\frac{\sigma}{E'+(2\pi\eta)/t}=\frac{\sigma t}{E't+2\pi\eta}. \tag{9a}$$

As long as $E''\ll E'$, the second term in the denominator can be neglected with respect to the first one and one obtains

$$\epsilon\simeq\sigma/E'. \tag{10}$$

This is the one limiting case of the Nutting equation corresponding to t^0 and $\psi\simeq 1/E'$. If $E''\gg E'$ the first term in the denominator can be neglected, one obtains

$$\epsilon=\sigma t/2\pi\eta. \tag{11}$$

This is the other limiting case of the Nutting equation

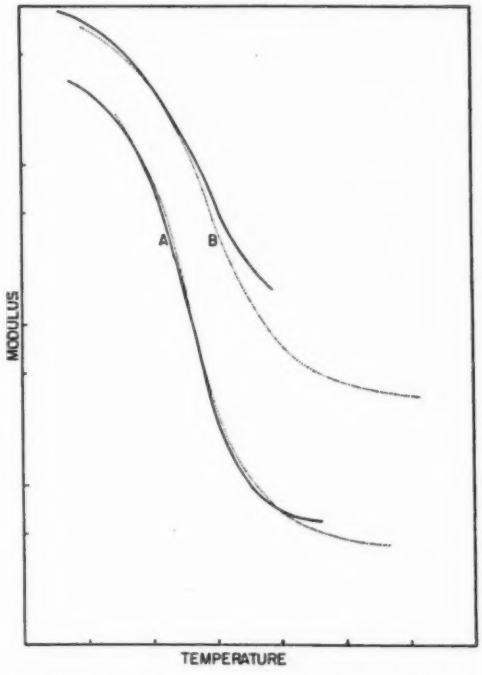


FIG. 6. Comparison between dynamic modulus and 5 second creep modulus.

corresponding to t^1 and $\psi\simeq 1/2\pi\eta$. In many cases—as for example in the transition region of all polymeric substances—E'' cannot be neglected with respect to E', or vice versa. One must, therefore, retain in the intermediate range, when n lies between 0 and 1, the complex modulus. The intensity of E^* is given by $|E|=(E'^2+E''^2)^{\frac{1}{2}}$ whereas the contribution of time (t) to $|E|$ is proportional to $\sin\phi$ which is equal to E''/E^* (see Fig. 7). It will be recognized that the limiting cases of $\sin\phi=\{{}^{0}_{1}\}$ correspond to the limiting cases of the Nutting equation. Equation (9) in terms of real quantities takes the form (always assuming $\beta=1$)

$$\epsilon\simeq\frac{\sigma t^{\sin\phi}}{(E'^2+E''^2)^{\frac{1}{2}}}=\frac{\sigma t^{E''/E^*}}{(E'^2+E''^2)^{\frac{1}{2}}}. \tag{12}$$

† The modulus scales in Fig. 6 are shifted slightly because the time scale in the two experiments is not the same; however, the shapes of the curves are almost identical over an extended temperature region.

[13] H. Leaderman, *Elastic and Creep Properties of Filamentous Materials* (Textile Foundation, Washington, 1943).

[14] Buchdahl, Nielsen, and Merz, J. Polymer Sci. **6**, 404 (1951).

[15] G. E. Scott Blair, *A Survey of General and Applied Rheology* (Pitman Publishing Corporation, New York, 1944), p. 145.

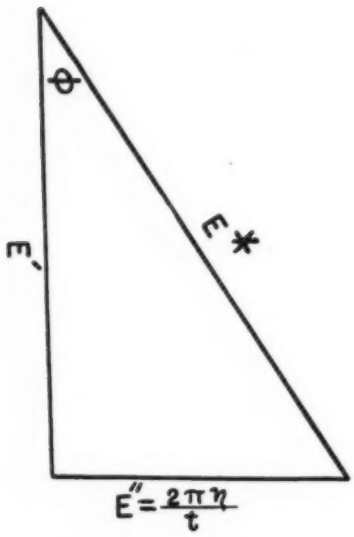


FIG. 7. Relationship between E^*, E', and E''.

As no objection can be raised against Eq. (9) on the basis of dimensional analysis, Eq. (12) should be dimensionally correct also since [time] is still partly tied up in the denominator to balance out the exponent in the numerator in either equation.

Comparing (12) with (2), it follows that

$$n \simeq E''/|E^*| \tag{13}$$

and

$$\psi \simeq \frac{1}{(E'^2+E''^2)^{\frac{1}{2}}}. \tag{14}$$

Two equations (8) and (13) have been given to relate "n" with dynamic mechanical quantities. There is a significant difference between the two expressions: $E''/(E^*)$ varies between 0 and 1 whereas E''/E' varies between 0 and ∞. Equation (8) should not be used to calculate n for large values of E''/E' because no experimental data are known where n becomes larger than one (it certainly never does approach infinity). Using $E''/(E^*)$ to calculate n does not give a good quantitative agreement with the data given in Table I and Fig. 5. In evaluating a large number of creep curves, it was found that n never exceeded 0.7 unless the deformation was predominantly a purely viscous one, $n \to 1$. It seems reasonable, therefore, to introduce a limiting upper value for the parameter n which in combination with $E''/(E^*)$ gives satisfactory agreement with experimental data.

$$\epsilon \simeq \frac{\sigma t^{2E''/\pi|E^*|}}{(E'^2+E''^2)^{\frac{1}{2}}}, \tag{15}$$

where $n=2E''/\pi|E^*|$. This expression for n is the same as derived above on the basis of Zener's equation except that E' is replaced by $|E^*|$.

THE COEFFICIENT β AND THE EFFECT OF STRESS ON n

The Nutting equation contains a coefficient β which so far has been assumed to be equal to one. In reality deviations from the value $\beta=1$ are quite frequent; for the polyvinyl chloride composition given above the average value of β is about 0.80 and does not seem to vary in a systematic manner with temperature. For polystyrene the compliance $\{\epsilon(t)/\sigma=J(t)\}$ is also a function of the stress (14). It is well known that in the study of flow curves ($d\epsilon/dt$=constant) one observes very frequently a shear dependence which is characteristic of non-newtonian fluids. In recent publications[16] Scheele and collaborators have tried to show that β goes through a maximum in the transition region similar to n; except that in the latter case β approaches a value of one at either side of the maximum. It is, however, questionable that creep data below the second-order transition temperature can be correctly represented by a constant creep rate in view of the data presented in this paper. (According to the Nutting equation $(d\epsilon/dt) \sim nt^{n-1}$ which implies a variable rate of shear unless $n \sim 1$; see also the paper by A. Smekal.[17]) The stress dependence of the creep rate has been studied by Findley[18] and Sauer[19] for polystyrene at temperatures well below the second order transition temperature. The results obtained by these investigators show that β varies between 0.5 and 1.3 depending on the stress; the problem is very similar in its complexity to phenomena of strain-hardening in metals.

Very little is known about the stress dependence of n or $E''/(E^*)$. Unpublished data obtained by one of us (L.N.) indicates that on the low temperature side of the transition the dissipation coefficient is rather insensitive to stress variations at low stress levels. However, in the transition region the dissipation factor increases with increasing stress rather markedly. One would, therefore, expect an analogous behavior for n. However, such data as are available are not precise enough to determine with certainty whether or not this is so, although on the basis of Sauer's data[19] at room temperature and tensile creep measurements in the transition region at large and small stresses it would appear that the analogy carries over.

THE PHYSICAL INTERPRETATION OF n

Although the details of the relationship between n and the dissipation factor will, most likely, require certain modifications, there can be no doubt about the basic interdependence of these quantities. Kuhn[20] and Tobolsky[21] showed, some time ago, that for materials

[16] Scheele and Timm, Kolloid Z. **116**, 129 (1950); Kolloid Z. **120**, 103 (1951).
[17] A. Smekal, Kolloid Z. **116**, 112 (1951).
[18] W. N. Findley, J. Appl. Phys. **21**, 258 (1950).
[19] Sauer, Marion, and Hsiao, J. Appl. Phys. **20**, 507 (1949).
[20] Kuhn, Kunzle, and Preissmann, Helv. Chem. Acta **30**, 307, 464 (1947).
[21] B. A. Dunell and A. V. Tobolsky, J. Chem. Phys. **17**, 1001 (1949).

in the rubbery state the creep and stress relaxation functions can be reasonably well approximated from dynamic mechanical data. The work of Zener and Kê,[13] in particular, has shown that a very good quantitative agreement can be obtained between measured and calculated stress relaxation and creep function in the case of metallic compounds. The results of this investigation show that this relationship between the essential parameter of the creep function (in terms of the Nutting equation) and the dissipation factor remains essentially the same whether a polymeric composition is a rigid high modulus solid, a soft low modulus rubber, or in between these two "states." It is therefore justified to postulate that for solids, in general, the creep function and the dissipation coefficient are manifestations of the same molecular processes. It is understood that transverse or intercrystalline thermal currents[13] which can make a very slight contribution to the dissipation factor will not be effective in the creep function. It follows then that the process or processes responsible for these phenomena should not be interpreted in terms of specific molecular models. For example, the creep function of polyvinyl chloride compositions is the same as that for platinum,[22] although the molecular structure of the two systems is certainly very different. Another example, along the same lines, is the interpretation of creep measurements of styrene-butadiene copolymers in terms of the molecular structure parameters as shown in Table II. A good qualitative correlation between n and the amount of permanent gel content (due to chemical crosslinking) is immediately apparent; n is influenced also by the semipermanent gel-content, as is shown by the data on polystyrene, but this is not measurable by swelling experiments. At first sight it appears reasonable to interpret n in terms of the gel structure of a rubber; however, rubber gels are certainly not the only molecular structures which give these particular creep functions. Therefore, the specific model of a rubber gel—a network consisting of carbon-carbon chains of essentially free rotation and held together at certain points by primary valence forces—should not be used as the basis of a theory of the time dependence of viscoelastic deformation.

[22] R. P. Carreber, J. Appl. Phys. **21**, 1289 (1950).

TABLE II. n and ψ as a function of molecular properties of styrene-butadiene copolymers.

Material	n	ψ (10 sec)	% gel	Swelling index	$[\eta]$ of sol fraction
1	0.600	9.4×10^{-7}	0		0.612
2	0.410	1.6×10^{-6}	0.5–1.0	140	1.530
3	0.432	1.6×10^{-6}	8.2	150	1.774
4	0.347	7.5×10^{-7}	66.6	65.8	0.639
5	0.350	1.7×10^{-6}	52.3	34.8	1.380
6	0.262	1.1×10^{-6}	78.5	27.4	0.328
7	0.242	8×10^{-7}	80.9	25.9	0.578

The dissipation factor measures the ratio of dissipation of energy to the total stored energy when a solid is subject to stress (or strain). The proportionality between the creep function and the dissipation factor means, therefore, that the creep function at any temperature is a measure of energy losses during a deformation process. The deformation of a solid is a cooperative phenomena, in the sense that it involves the simultaneous displacement of a large number of molecules and the degree to which this process takes place as a cooperative phenomena is reflected by the value n or $E''/|E^*|$. The upper limit of n corresponds to a deformation process where all the stored energy is being dissipated into kinetic energy of individual molecules, whereas the lower limit would correspond to a deformation process where no dissipation of energy is possible (perfectly ideal crystal). The parameter "ψ" measures the intensity of the deformation and is therefore primarily determined by the configuration and forces between the molecular units of the material. The dissipation factor $E''/|E^*|$ or n, on the other hand, depends primarily on the state of aggregation, i.e., the size and interaction of domains or clusters—which might or might not be crystalline. A comparison of the temperature dependence of the two parameters—ψ and n—illustrates the point. ψ is either some exponential function of T (when $n\rightarrow1$) or only very slightly dependent on T (when $n\rightarrow0$), whereas the temperature dependence of n is entirely different. It is in a transition region that n changes markedly with temperature; i.e., in regions where the state of aggregation undergoes a profound change.

Formation of Annealing Twins During Grain Growth*

R. L. Fullman and J. C. Fisher
General Electric Research Laboratory, Schenectady, New York
(Received May 11, 1951)

It is proposed that annealing twins form during grain growth as a result of a decrease in the interfacial free energy of grain boundaries that would not be achieved in the absence of twinning. Experiments designed to permit detection of the changes in grain boundary free energy accompanying the formation of twins give results in agreement with this concept. The abundance of twins found in annealed copper depends on the texture in a manner that is consistent with the theory.

INTRODUCTION

A DOMINANT feature of the microstructures of most annealed face-centered cubic metals is the presence of many straight-sided crystals that have a twinned orientation relative to their neighbors. These crystals are called annealing twins. A better understanding of the way in which they form during annealing should be of importance in understanding the origin of annealing textures in face-centered cubic metals and alloys.

In the absence of strain energy, the volume free energy of an annealing twin is the same as that of the original crystal, since they differ only in orientation. The free energies of the interfaces separating twinned crystals from each other and from neighboring crystals then are the only free energies concerned with the formation of annealing twins. When strain energy is present, as in recrystallization, the interfacial free energies still are important, but thermodynamic analysis is more complicated by lack of knowledge concerning the distribution of strain energy in deformed crystals.

The observations of annealing twins in metals have been reviewed extensively by Carpenter and Tamura[1] and by Mathewson.[2] Hull[3] pointed out that annealing twins may be found in most face-centered cubic metals, and subsequent investigations suggest that all form annealing twins after appropriate treatments, although the number formed may vary greatly. Aluminum is of particular interest in this respect, since it forms relatively few annealing twins.

Two processes whereby annealing twins might form have been described. Mathewson[2,4] suggested that annealing twins form by lateral growth of extremely thin deformation twins during annealing of face-centered cubic crystals. Photomicrographs presented by Burke and Shiau[5] suggest that lateral migration of twin boundaries in face-centered cubic metals may occur when strain energy is still important during the early stages of annealing. An alternate process has been described by Burke.[6] He observed that the annealing twins that formed during grain growth of alpha-brass appeared, one twin boundary at a time, at the corners of growing grains. During continued grain growth the twin boundaries were not observed to migrate. The theory of twin formation advanced by Mathewson[2,4]

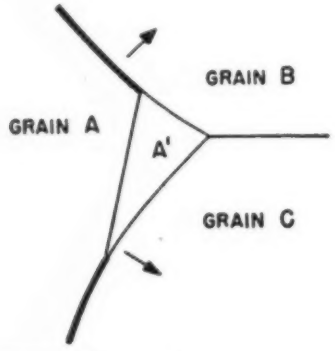


FIG. 1. Schematic representation of the changes in grain boundary free energy associated with the formation of a twin A' in a small region in the corner of a growing grain A. Grain boundaries with high interfacial free energies are represented by wide lines.

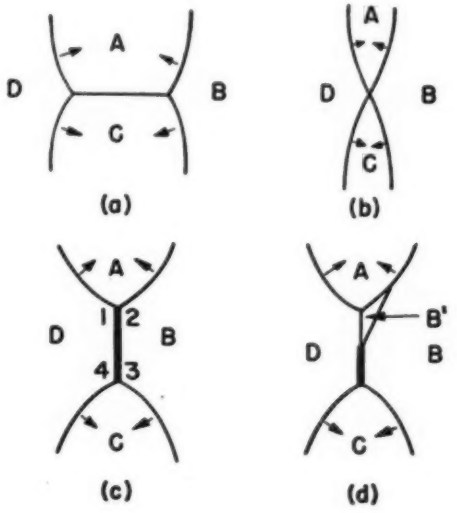


FIG. 2. Schematic representation of the disappearance of a grain boundary, illustrating how a change in the identity of a grain's neighbors could lead to the formation of an annealing twin in the corner of a growing grain.

* This paper is based upon a part of a dissertation submitted by R. L. Fullman to the faculty of the Engineering School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Engineering, June, 1950.

[1] H. C. H. Carpenter and S. Tamura, Proc. Roy. Soc. (London) **113A**, 161 (1926).

[2] C. H. Mathewson, Proc. Inst. Metals Division, Am. Inst. Mining Met. Engrs. **1928**, 7 (1928).

[3] A. W. Hull, J. Franklin Inst. **193**, 189 (1922).

[4] C. H. Mathewson, Trans. Am. Soc. Metals **32**, 38 (1944).

[5] J. E. Burke and Y. G. Shiau, Trans. Am. Inst. Mining Met. Engrs. **175**, 141 (1940).

[6] J. E. Burke, Trans. Am. Inst. Mining Engrs. **188**, 1324 (1950).

does not appear to account for the formation of annealing twins during grain growth after the recrystallization process is complete, for in face-centered cubic metals, migration of twin boundaries is not observed during grain growth after the strain energy responsible for recrystallization has largely disappeared.

PROPOSED THEORY

It is suggested that twin boundaries form during grain growth when the free energy of the boundaries between a grain's neighbors and its twin would be less than that of the boundaries between the neighbors and the grain itself. A twin boundary might then form in a corner of a growing grain and extend as the grain continues to grow as a twin of the original crystal. The conditions required for a decrease in free energy to accompany the formation of a twin are clarified by Fig. 1, which represents schematically the small region where atoms are being added to a growing grain A, and removed from the neighboring grains B and C. The small arrows indicate the directions of grain boundary migration, and the wide lines indicate boundaries of relatively high interfacial free energy. It may be seen that the free energy will be decreased by the formation of the twin A if:

$$s_{A'B}\sigma_{A'B}+s_{A'C}\sigma_{A'C}+s_t\sigma_t<s_{A'B}\sigma_{AB}+s_{A'C}\sigma_{AC}, \quad (1)$$

where s and σ represent the area and interfacial free energy, respectively, of the boundaries indicated by the subscripts. The subscript t refers to a coherent twin boundary, i.e., to a twin boundary that coincides with the twinning plane. In effect, formation of the twin boundary A/A' will cause an over-all free energy decrease if the A'/B and A'/C grain boundary energies are decreased over the A/B and A/C boundary energies by more than the A/A' twin boundary energy. Annealing twins will form in grain corners during grain growth when the geometry and orientation relationships are such that inequality 1 is satisfied. In copper, for example, the ratio of the interfacial free energy of coherent twin boundaries to that of average grain boundaries is only 0.035,[7] much less than the variability of grain boundary free energy, so that twin boundary formation should be frequent.

The identity of a grain's neighbors may change during grain growth when the boundary between two adjacent grains disappears, as in Fig. 2(a) to (c). An example may be found in Burke's[8] photographs of grain boundary migration in zinc. If the new grain boundary between grains B and D has relatively high interfacial free energy, as indicated, there are four new corners (numbered on the figure) at which conditions favoring the formation of a twin, as in Fig. 2(d), may exist.

A change in the identity of a grain's neighbors may also occur as a result of interaction with existing twins. Figure 3 shows a grain A growing at the expense of grains B and C, one of which contains a twin B'. If the interfacial free energy of the A/B' grain boundary is relatively high, and the geometry is favorable, a twin A' will form in the advancing corner of grain A when the configuration of Fig. 3(b) is reached. When the boundary has advanced to the second previously existing twin boundary in grain B, the advancing grain A' may twin back to the original orientation A, as in Fig. 3(d). Figure 4 shows an alternative sequence that may occur if the free energy of an A'/B boundary is high, since further progress of the A/B boundary, as in Fig. 3, may then be prevented. Even when the twin boundary planes are not aligned so that they can coincide along a line where they meet the grain boundary, a minimum free energy position for the A/B boundary exists, leading to a tendency to retard migration of the boundary. Twin boundaries near grain corners of adjacent grains very frequently meet the grain boundary at a common point.

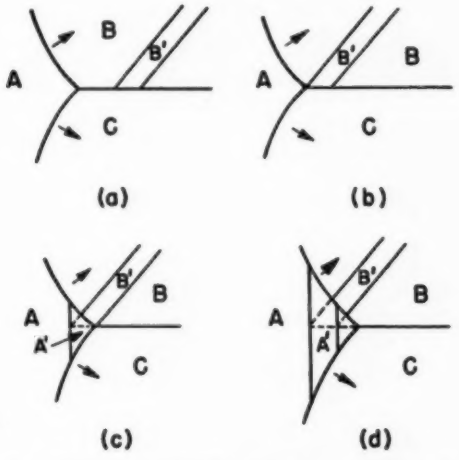


FIG. 3. Schematic representation of the formation of a twin in the corner of a growing grain, as the result of a change in the identity of the grain's neighbor when the advancing grain boundary meets an existing twin.

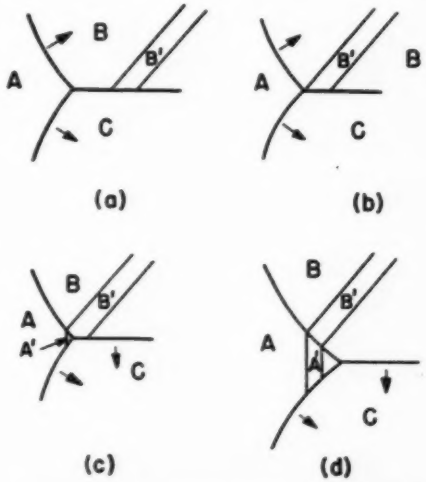


FIG. 4. Schematic representation of a possible influence of twin formation on grain boundary migration. Association of the twins A' and B' may retard migration of the A/B grain boundary.

[7] R. L. Fullman, J. Appl. Phys. **22**, 448 (1951).

[8] J. E. Burke, "Grain control in industrial metallurgy," Am. Soc. Metals, Cleveland, Ohio (1949).

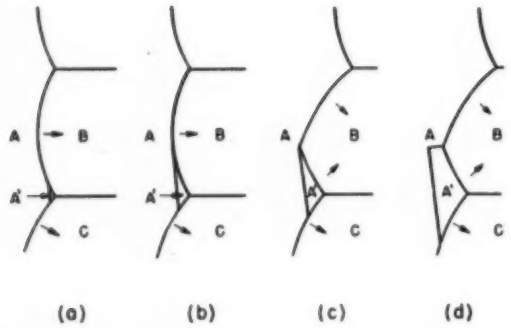


FIG. 5. Schematic representation of a way in which a noncoherent twin boundary might form, after formation of a twin in a corner of a growing grain.

Steps of noncoherent twin boundary[9] along the side of a coherent twin boundary and twin bands that end within the grain could result from the sequence of boundary configurations shown schematically in Fig. 5. Here a coherent twin boundary that formed in the corner of grain A has become tangent to the A/B grain boundary upon migration to the configuration of Fig. 5(b). Continued migration of the grain boundaries must lead to formation of a noncoherent twin boundary segment, as shown in Fig. 5(c) and 5(d). If a second twin boundary, parallel to the first, forms in the advancing corner of A', continued grain boundary migration will cause the new coherent twin boundary to extend until it meets the noncoherent twin boundary. The result is a twin A' that ends within grain A. On the other hand, if a second twin boundary forms in the corner of grain A above the corner where the first twin boundary formed, continued grain boundary migration results in the formation of a "stepped" twin boundary.

If the free energy of a grain boundary at the end of a twin is sufficiently higher than that of the adjacent boundary, the twin may pull away as shown schematically in Fig. 6. The process is accompanied by a free energy decrease if:

$$s_{A'B}\sigma_{AB}+s_{t'}\sigma_{t'}<s_{A'B}\sigma_{A'B}+s_{t}\sigma_{t}, \qquad (2)$$

where the subscript t' refers to a noncoherent twin boundary. A twin that ends within a grain may disappear by migration of the noncoherent boundary at its end, with a decrease in the coherent twin boundary area, as shown in Fig. 7. If a whole series of fine parallel twins pulled away from a grain boundary in the manner shown in Fig. 6 and disappeared, the result would be similar to that described by Mathewson[2] as "coalescence" of twins, a mechanism for the apparent broadening of twins during grain growth.

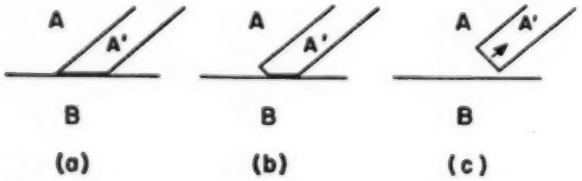


FIG. 6. Schematic representation of the dissociation of a twin from a grain boundary, leading to disappearance of the twin.

[9] R. L. Fullman, J. Appl. Phys. 22, 456 (1951).

In addition to the simple formation of a single twin boundary in a grain corner as illustrated in Fig. 2, more complex behavior may be expected, as when several of the new corners form twins on continued growth. There also may be changes in the general direction of boundary migration as a result of the changes in relative interfacial free energies resulting from formation of a twin. Only a two-dimensional analysis of relatively simple situations has been attempted in the present treatment.

EXPERIMENTAL VERIFICATION

Since the intervals of distance between changes in the identity of a grain's neighbors should be roughly proportional to the grain size of the material being consumed, it follows from the proposed theory of annealing twin formation that the spacing between twin boundaries should also be proportional to the grain size being consumed. This prediction is verified, for during normal grain growth (no marked contrast in the grain sizes) the spacing between twin boundaries is roughly proportional to the grain size, and hence to the grain size of the material being consumed. Similarly, when exaggerated grain growth or secondary recrystallization has occurred, and a very large grain has grown at the expense of a fine-grained matrix, it is frequently observed that the large grain contains many very fine twins, whose size is comparable with the grain size of the material being consumed.

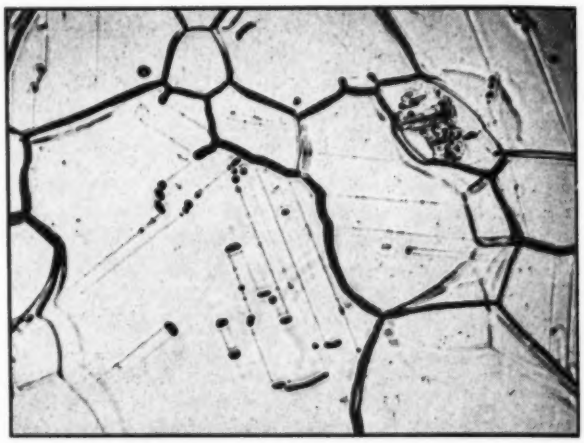


FIG. 7. Surface of compression-rolled copper sheet thermally etched for 100 hours at 945°C in an atmosphere of lead vapor and hydrogen. Rounded grooves indicate previous positions of noncoherent twin boundaries that have migrated in the directions required for a decrease in area of the coherent twin boundaries. 50×.

A direct test of the validity of the proposed mechanism of twin formation during grain growth would be

provided by observations of the relative interfacial free energies of grain boundary segments on opposite sides of twin boundaries that are newly formed in grain corners, and hence still have as neighboring grains those that were present when the twinning operation occurred. If the theory proposed is correct, the average free energy of the grain boundary segments toward the corner should be smaller than on the side of the twin boundary away from the corner.

If the B/C grain boundary of Fig. 8 has average interfacial free energy, and the A'/B and A'/C grain boundaries have less than average free energy, balance of interfacial tensions would lead to the angle θ having a value less than 120°. The significance is not reduced if some of the twins selected for observation are being consumed, rather than being formed, for grains that are being consumed are essentially random and should contribute values of θ averaging 120°.

Specimens of OFHC copper bus-bar 0.200 inch in thickness were annealed at 500°C for one hour and compression rolled† in accordance with one of two rolling schedules as follows: (1) Compression roll to

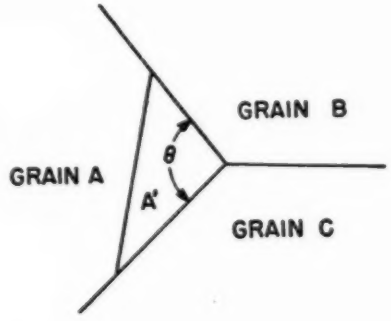


FIG. 8. Diagram showing the angle θ measured in twinned grain corners.

0.020-in. thickness, with two intermediate anneals at 500°C. (2) Compression roll to 0.025-in. thickness, with no intermediate anneals. Specimens prepared by either of these schedules were annealed for 18 hours at 1040°C. Observations were made on the angles, θ, formed between grain boundary traces in twinned corners for which the distance from the grain corner to the twin boundary was less than 10 percent of the grain diameter, and no twin boundary in either adjacent grain was so close to the corner. Measurements were made at 148 grain boundary intersections meeting these requirements. No significant difference appeared between the measurements made in specimens prepared by the two rolling schedules used. A bar diagram of the distribution observed on all samples is shown in Fig. 9. The mean value of θ found was $115.20 \pm 0.9°$. According to standard statistical formulae, the probability that the mean value of θ is actually less than 120° is 0.99974.

The results of the experiment just described indicate that the grain boundary free energy toward a grain corner from a newly formed annealing twin boundary is

† In compression rolling the rolling direction is changed at each pass of the specimen through the rolls.

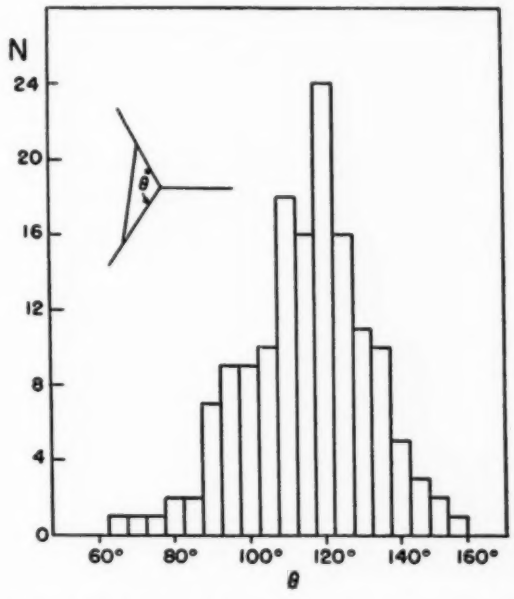


FIG. 9. Distribution of angles (θ) measured at twinned grain corners on the surface of annealed copper sheet.

smaller than the grain boundary free energy on the opposite side of the twin boundary. The observation was checked by means of deep thermal etching in lead vapor. The relative widths of the grooves visible on examination of a thermally etched surface provide a qualitative measure of the relative interfacial free energies of the boundaries responsible for the formation of the grooves.

A strip of OFHC copper, compression rolled in accordance with schedule (2) described above, was mechanically polished and then thermally etched for 16 hours at 900°C in lead vapor and a partial atmosphere of hydrogen. Grain corners containing a twin boundary within 10 percent of the grain diameter from the corner of one grain, and no other twin boundaries closer to the corner of any of the three grains, were selected for

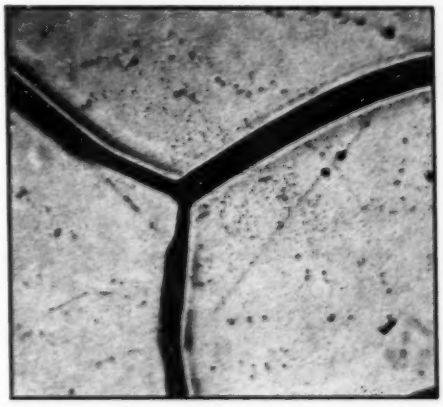


FIG. 10. Example of a change of type *Narrower, Narrower* in the apparent width of grain boundary grooves at a twinned grain corner. Copper sheet thermally etched in lead vapor for 10 hours at 945°C. 100×.

TABLE I. Changes in grain boundary grooves at twinned grain corners in thermally etched copper.

Type of change in groove width	Number of observations by Author	Independent observer	Total	
Narrower, Narrower*	10	3	13	consistent with theory
Narrower, Same	38	52	90	
Narrower, Wider	5	9	14	indeterminate
Same, Same	17	10	27	
Same, Wider	12	5	17	inconsistent with theory
Wider, Wider	3	3	6	

* Groove narrower toward corner for grain boundaries at both ends of twin.

examination at a visual magnification of approximately 235×. Figure 10 shows an example of a change in grain boundary groove width. A boundary groove that appeared narrower toward the grain corner than on the side of the twin boundary away from the grain corner was denoted as *Narrower*, while a boundary that appeared wider toward the grain corner was denoted by *Wider*. Boundaries for which no difference in width of the groove could be positively established, or for which a small apparent difference in width could be reversed by changing the focusing position of the microscope, were called *Same*. Observations of all the suitable twinned grain corners found on the specimen were made by two individuals, one an independent observer who knew nothing of the theory of twin formation that has been described. The results are summarized in Table I.

Of 167 observations of grain boundary grooves adjacent to twins in grain corners, 103 were consistent with the proposed theory, 41 were in doubt, and 23 were inconsistent. If the proposed theory is correct, the observations that are not consistent with the theory must be attributed to the presence of twins that are being consumed. The probability of *Same* observations should be identical for new and old twins, since these observations reflect primarily the sensitivity of the observations. If the probabilities of *Narrower* and *Wider* observations are equal for the twins being consumed, i.e., if these twins make a random contribution to the distribution, then the observed distribution is consistent with the proposed theory with 56 percent of the twins observed being new ones. The probability that a randomly chosen grain corner is moving toward or away from a nearby line is $\frac{1}{2}$, so that to a first approximation 50 percent of the twins should be new ones. The slight preferential selection of newly formed twins is consistent with the fact that large grains are more likely to be growing than small grains[8] and the existence of a twin boundary within 10 percent of the grain diameter from a corner is more probable in large grains than small ones.

THE ABUNDANCE OF TWINS

No precise measurements of the abundance of twins over a wide range of grain sizes have been reported. Smith[10] pointed out that the number of twins per grain is essentially constant during grain growth. Hibbard, Liu, and Reiter[11] counted the twins in copper and alpha-brass rolled and annealed at various temperatures. In copper they found a nearly constant number of twins per grain, but the grain size varied by less than a factor of two. In brass the number of twins per grain decreased from about 7 to 2 with a fourfold increase in grain size.

According to the theory proposed, the abundance of twins should decrease with increasing ratio of twin boundary free energy to grain boundary free energy, hence with increasing perfection of preferred orientation, which leads to grain boundaries with smaller interfacial free energy. It is well known that copper with a very perfect "cube" texture (produced by annealing after severe rolling) has almost no twins, while Hibbard, Liu, and Reiter[11] found about 1.6 twin boundaries per grain‡ in rolled and annealed copper that did not develop the cube texture. In order to extend this correlation to a less perfect texture, the abundance of twin boundaries was measured in compression rolled copper sheet after annealing.

The samples were prepared in such a way that two widely different grain sizes would be obtained, with grains extending completely through the sheet for both grain sizes. Two strips of OFHC copper were compression rolled in accordance with schedule (1) described above (final thickness 0.020 in.) and annealed for 18 hours at 1040°C in hydrogen, leading to an average grain diameter of 0.90 mm. Another specimen was compression rolled from 0.200-in. to 0.006-in. thickness with no intermediate anneals, mechanically polished to a thickness of 0.005 in., and annealed for 1 hour at 800°C in hydrogen, leading to an average grain diameter of 0.21 mm. The samples were examined on the metallograph at a visual magnification of approximately 100×, using polarized illumination to assist in the identification of twins and grain boundaries.

The number of twin boundaries per grain found in the compression rolled copper with an average grain diameter of 0.21 mm was 5.08 ± 0.18. In the copper with a grain-size of 0.90 mm the mean number of twin boundaries per grain was 4.58 ± 0.15. The abundance of twins was nearly the same for the two grain sizes examined in compression rolled copper, and was substantially larger than that found by Hibbard, Liu, and Reiter[11] in straight rolled copper. Conditions that lead to less random orientations of neighboring grains seem to lead to a smaller number of twins per grain, as would be predicted on the basis of the suggested theory of twin formation and disappearance.

[10] C. S. Smith, Trans. Am. Inst. Mining Met. Engrs. **175**, 15 (1948).

[11] Hibbard, Jr., Liu, and Reiter, Trans. Am. Inst. Mining Met. Engrs. **185**, 635 (1949).

‡ Hibbard, Liu, and Reiter counted the number of completely enclosed "twins" per grain. The number of twin boundaries per grain may be from one to two times the number of twins counted in this manner, or 1.06 to 2.12.

CONCLUSIONS

Direct observations of the formation of annealing twins at moving grain corners, of the proportionality of twin boundary spacing to the grain size being consumed during growth, of the significantly small average grain boundary dihedral angle in twinned grain corners, of the relatively shallower thermally etched grain boundary grooves on the grain corner sides of newly formed twins, and of the dependence of twin abundance on texture, all support the proposal that the presence of annealing twins at the corners of certain grains corresponds to a lower total interfacial free energy than would the absence of twins, and that new stationary annealing twin boundaries appear at moving grain corners whenever a decrease in the over-all interfacial free energy would result.

ACKNOWLEDGMENT

The authors are indebted to many members of the Metallurgy Division of the General Electric Research Laboratory for assistance, and in particular to G. W. Sears. Professor F. C. Frank, of the University of Bristol, read and criticized the manuscript. The investigation was supported by the U. S. Atomic Energy Commission.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

Discrete Energy Levels of a Nonlinear Oscillator

JAMES H. MORAN
Department of Electrical Engineering, Case Institute of Technology, Cleveland, Ohio
(Received May 15, 1951)

Following a brief survey of the theory, especially Eckweiler's work, of nonlinear oscillators satisfying an equation of the form $(d^2q)/(dt^2)+\epsilon F(dq/dt)+q=0$ a special case is considered for which the stable limit cycles have a definite relation to one another and to the so called Planck oscillator. A nonlinear "quantum effect" is discussed and shown to be similar to the harmonic quantum oscillator in some respects. No physical model is proposed for which the equation considered would be a description but various possibilities in this regard are indicated.

I. INTRODUCTION

IN Planck's original development of the quantum theory the linear harmonic oscillator played a central role. In order to obtain a formula in agreement with experiments on blackbody radiation Planck was led to the assumption that his elementary oscillating systems would attain only a discrete set of energy levels instead of the continuous range generally associated with a system behaving according to the laws of classical mechanics. Because some mechanism was called for to allow for the possibility of a change in the system configuration he further assumed that the elementary oscillators could alter their energy in a discontinuous manner. No attempt was made to follow the system through these "jumps" or transitions from one energy level to another.

These ideas of Planck's have a close analogy in the mechanics of nonlinear systems which will be indicated in this paper.

II. THE HARMONIC OSCILLATOR

The equation of a linear harmonic oscillator of natural frequency f is in terms of a position coordinate q

$$(d^2q)/(dt^2)+4\pi^2f^2q=0. \tag{1}$$

The frequency f depends of course on the "mass" of the oscillator and its "spring constant." If we introduce the momentum $p=mdq/dt$ the energy of the oscillator is given by

$$E=p^2/2m+4\pi^2f^2mq^2. \tag{2}$$

The solution of (1) is often represented as a curve drawn in the p, q plane—the so called phase plane. A curve of constant energy in this plane is from (2) an ellipse whose semi-axis along p and q are respectively

$$a=(2mE)^{\frac{1}{2}};\quad b=(2E/(4\pi^2f^2m))^{\frac{1}{2}}. \tag{3}$$

The area of such an ellipse is $\pi ab=E/f$. Planck's assumption amounts to the statement—the oscillator will find itself only in such states that its representative point in the phase plane lies on one of a discrete set of ellipses of the above type where the set is such that adjacent members differ in area by a constant value. This constant difference in area is a new universal constant, Planck's constant. It is this particular aspect of Planck's work that has such a close analogy with certain nonlinear systems.

III. THE NONLINEAR OSCILLATOR

The idea of a system capable of only discrete energy levels is not compatible with classical mechanics. However if we agree to neglect transient conditions during a change from a motion specified by one energy level to another and require only that the possible steady states or equilibrium states be associated with

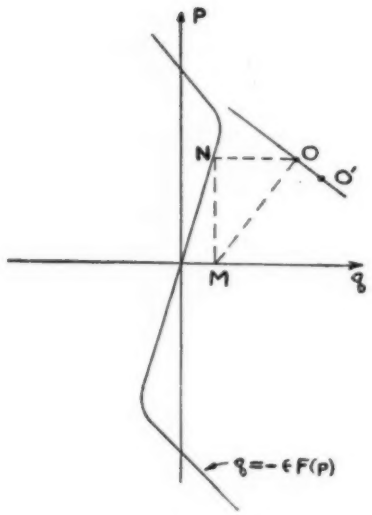


FIG. 1. Graphical construction for Eq. (6).

Planck's conditions then a suitable nonlinear system will fulfill the requirements. In any event transient states in an atomic oscillator must be of extremely short duration.

In order to make the main idea clear, unclouded by a mass of symbolism, let us study a system in which we have numerically $m=2\pi f=1$. We will also take Planck's area difference to be a unit area.

We will make the important assumption that the system operates with a slight nonlinearity which depends on the velocity $dq/dt=\dot{q}$ so that the equation is

$$d\dot{q}/dt+\epsilon F(\dot{q})+q=0. \tag{4}$$

Where we assume further that ϵ is small but not zero and that $F(\dot{q})$ denotes a bounded function of the velocity for the moment unspecified.

Equation (4) has received considerable attention from workers in the field of nonlinear mechanics. In many cases, depending on the form of $F(\dot{q})$, the solutions are known to be oscillatory and to have the property of approaching harmonic oscillations as ϵ becomes indefinitely small. As a matter of fact a typical solution $q(t, \epsilon)$ as a function of ϵ can be shown to approach a harmonic oscillation in the limit $\epsilon\to 0$, $t\to\infty$. It is found, however, that the harmonic oscillation so obtained has a definite amplitude or set of amplitudes. This fact is of primary importance for this discussion. A very well-known example of such behavior occurs when $F(\dot{q})=\dot{q}-1/3\dot{q}^3$ in which case the equation is that first studied by Van DerPol. Here there is a single amplitude of the above type which is in fact known to be 2. This result means that for ϵ very small and $F(\dot{q})$ of the above form the steady state is a continuous oscillation $q=2\cos t$ independent of the particular initial conditions from which the system starts. In the phase plane this solution is a curve of constant radius $r=2$, i.e., a circle.

Equation (4) is generally studied most readily through a graphical approach. The time variable can be eliminated if we introduce the momentum $p=dq/dt$ so that (4) becomes

$$dp/dt+\epsilon F(p)=-q. \tag{5}$$

The time variable can then be eliminated to give

$$dp/dq=(-\epsilon F(p)-q)/p. \tag{6}$$

When $F(p)$ is given either in the form of a curve or an analytic function, Eq. (6) lends itself to immediate graphical solution. The right-hand side of (6) can be evaluated graphically in a simple manner for any specified value of p and q. Thus for every point of the phase plane we are able to determine the slope of the solution curve. Evidently by choosing successive points along the indicated slopes close to one another we can trace out a solution curve. The curve determined in this way will, of course, depend on the starting point, i.e., the initial conditions. For functions $F(p)$ of a certain type, however, it is easily seen that there are a set of curves, the so-called limit cycles, which every solution curve approaches, after a transient period, independently of the initial conditions. The only effect of the initial conditions is to select a particular member of the set of limit cycles. In the case of the Van DerPol oscillator there is only one member in the set so that the solution curve always approaches a single limit curve independently of the initial conditions.

The details of the graphical solution are easily understood by reference to Fig. 1 which is a drawing of the phase plane. The first step is to plot on this plane the equation $q=-\epsilon F(p)$. A possible case is shown. Given initial conditions correspond to a point O in the plane. Suppose the system is in a state corresponding to O; then the slope of the solution curve through O is found as follows. Draw a line from O parallel to q until it intersects the curve $q=-\epsilon F(p)$. Drop a perpendicular from this point N to the q-axis at M. The line MO thus has a slope given by $p/[q+\epsilon F(p)]$. Thus by using M as center and constructing a perpendicular to MO through O the line OO' has the slope $[-q-\epsilon F(p)]/p$ which is the required slope through O and thus a new point O' can be determined. The same procedure may be used at O' to determine O'', etc., and the complete solution curve may be drawn.

This graphical analysis has been applied by various authors to a wide variety of problems, differing from one another in the choice of $F(p)$. It has been found from these studies that generally if $F(p)$ is an odd function of p a set of limit cycles are determined. Further, these limit cycles have the form of circles in the pq plane for very small ϵ values. The function appearing in the theory of the Van DerPol oscillator leads to a single limit cycle while a function such as $F(p)=\sin p$ leads to a infinite number of limit cycles. The number of limit cycles and their separation are determined by the particular form given to $F(p)$. Especially significant is the number and location of the zeros of $F(p)$.

IV. THE PLANCK OSCILLATOR

Eckweiler[1] has studied the relation between the form of $F(p)$ and the amplitudes of the limit cycles for the special case of vanishing ϵ. In the phase plane these amplitudes are apparent as the radii of the circles which are the limit cycles in this case. More particularly his analysis is concerned with the so-called "bifurcation" amplitudes which are the limiting values as ϵ approaches zero. We will review Eckweiler's work briefly so that we can see how an $F(p)$ of the required form might be determined. The requirement of $F(p)$ is that it have among its "bifurcation" amplitudes those chosen by Planck. If we denote these amplitudes by A_n then Planck's requirement is that the area difference between successive allowed cycles be unity. A detailed study of the stability of the limit cycles shows there to be an unstable cycle separating every pair of stable cycles. Now of course it is the stable cycles that are of interest to us while Eckweiler's analysis gives "bifurcation" amplitudes for both stable and unstable cycles. To remove this difficulty we will assume that the first and every odd-numbered cycle is stable while the even numbered cycles are unstable. Planck's condition is then $\pi A_{n+2}^2-\pi A_n^2=1$ where n is odd. We could assign the even numbered amplitudes in any fashion as long as they separate the odd-numbered ones. We will suppose they also satisfy $\pi A_{n+2}^2-\pi A_n^2=1$ so that in general this relation holds for the bifurcation amplitudes. To completely determine the A_n we will then take $A_0^2=0$ and $A_1^2=1/2\pi$. Then it follows that in general $A_n^2=n/2\pi$.

Eckweiler assumes a set of periodic solutions of (4) to exist for sufficiently small ϵ. We label these $q_n(t, \epsilon)$. These solutions are assumed to be continuous in ϵ so that

$$\lim_{\epsilon\to 0} q_n(t, \epsilon)=q_n(t, 0)=A_n \cos t. \tag{7}$$

Further the frequency $f(\epsilon)$ which in general depends upon ϵ is assumed to be a differentiable function of ϵ so that denoting $2\pi f$ by ω

$$\left.\frac{d\omega}{d\epsilon}\right]_{\epsilon=0}=\lim_{\epsilon\to 0}\frac{\omega(\epsilon)-1}{\epsilon}. \tag{8}$$

With the aid of these assumptions Eckweiler derives an integral equation for the function $F(p)$. We will assume $F(p)$ to be an odd function of p although this is not necessary. Then if we define $G(A)$ by

$$G(A)=\int_0^{2\pi} F(A \sin t) \sin t\, dt \tag{9}$$

the bifurcation amplitudes are determined from $G(A_n)=0$. Since the A_n are specified in our problem a function $G(A)$ can be immediately constructed for which $G(A_n)=0$. Once $G(A)$ is specified (9) may be considered an integral equation for $F(p)$. Through a change of variable[2] (9) can be converted into Abel's integral equation so that an explicit solution may be obtained. Unfortunately the form $F(p)$ takes is quite complicated and adds little to this discussion. The details of this calculation are given in the Appendix.

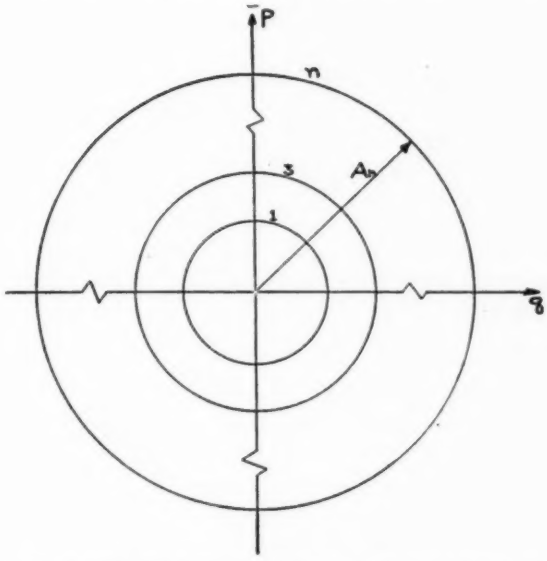


FIG. 2. Limiting form of the stable limit cycles for $\epsilon\to 0$.

V. DISCUSSION

The general features of the oscillator with the required $F(p)$ are illustrated in Fig. 2 where a few of the limit cycles are shown on the pq phase plane. The system will normally be in a state corresponding to one of these limit cycles where the amplitude of the oscillation is represented by the radius of the corresponding circle and the energy is proportional to the area of the circle. We can visualize the systems representative point moving from one curve to another due to a sudden impulse caused by an interaction with some external source of energy. If we observe the system only in its steady states the change in energy accompanying a change in state will be a definite quantum jump. Furthermore, an external agency with energy below a critical amount will be incapable of changing the state of the system.

In the steady state the nonlinear term will contribute no net energy to the system. Even during a transition under an impulsive force it will not contribute anything. It makes its effect felt only when the system finds itself in a condition not corresponding to one of the limit cycles. Under such a condition it judiciously adds to or subtracts from the energy of the system until it attains a limit cycle. The type of nonlinearity chosen evidently calls for some outside source of energy during these

[1] Eckweiler, Flanders, Friedrichs, John, and Stoker, *Studies in Non-Linear Vibration Theory* (New York University, 1946).

[2] See Appendix for details.

adjustments. The question of what model of the physical system would lead to such a nonlinearity is of course the important but difficult one.

In developing a suitable explanation for the nonlinearity there are at least two different viewpoints we might adopt. We might avoid a mechanical model by attempting to introduce a slight nonlinearity into Maxwell's field equations. It can be shown that the solution of Maxwell's equations for the problem of blackbody radiation is mathematically equivalent to an infinite number of harmonic oscillators. If a nonlinearity could be introduced into the field equations such that each effective oscillator of the infinite set could be treated as above the radiation formula of Planck could be derived without further hypothesis. The amplitudes of the various oscillators or modes would be restrained in the steady state to a discrete set of values. In this case it appears that the nonlinear term would have to be a function of the rate of change of the electric field intensity instead of the momentum.

On the other hand we might seek the nonlinearity in a mechanical model. This approach seems to require the moving electron to be part of the mechanical system. That the damping depends upon the velocity as a first order correction can be justified on the basis of classical electrodynamics.[3] Actually the damping is shown to be proportional to the negative of the rate of acceleration but in a system that satisfies Eq. (4) it is evident that $q=-\ddot{q}$ to a first approximation and hence $\dot{q}=-\dddot{q}$ so that we see the damping term may as well be considered proportional to the velocity. On the other hand there is no suggestion in the classical theory of the dependence on velocity required here. This would, it appears, only result upon rather special assumptions as to the forces which act on an electron. These forces, velocity dependent, of course would have to lead to negative damping terms in the equation of motion reminiscent of similar factors occurring in the theory of the maintenance of oscillations in vacuum tube circuits.

All of the above remarks are of course merely speculation. The basic idea that effects seemingly peculiar to quantum mechanics might be explained on the basis of some slight nonlinearity in the classical description has been considered quite recently by Hartley[4] and earlier by Born.[5] (They both deal with modifications of Maxwell's field equations.) The idea has a philosophical appeal but working out the details would obviously lead to great mathematical as well as physical difficulties. However it is interesting to conjecture what progress could be made in this regard with the help of but a fraction of the effort devoted to quantum mechanics in the last fifty years.

[3] H. A. Lorentz, *The Theory of Electrons* (J. A. Barth, Leipzig, 1909).

APPENDIX. THE CALCULATION OF $F(p)$

We follow Eckweiler's method in the following. The investigation begins with Eq. (9). Certain changes of variable are made.

$$\xi=A^2; \qquad \sigma=\xi\sin^2 t;$$
$$F(\sigma^{\frac{1}{2}})=H(\sigma); \quad g(\xi)=(2AG(A))/\pi^{\frac{1}{2}}. \tag{1}$$

Then Eq. (9) takes the form

$$g(\xi)=\pi^{-\frac{1}{2}}\int_0^{\xi} H(\sigma)(\xi-\sigma)^{-\frac{1}{2}}d\sigma. \tag{2}$$

The solution of Eq. (2) is well known

$$H(\xi)=\pi^{-1}\int_0^{\xi}\frac{dg(\sigma)}{d\sigma}(\xi-\sigma)^{-\frac{1}{2}}d\sigma. \tag{3}$$

Now $g(\xi)$ may be immediately constructed for our problem since we see from its definition that it must vanish for those values of ξ; $\xi_n=A_n^2=n/2\pi$. A suitable choice then is $g(\xi)=\sin 2\pi^2\xi$. Using this in Eq. (3) we get

$$H(\xi)=2\pi^{\frac{1}{2}}\int_0^{\xi}\cos(2\pi^2\sigma)(\xi-\sigma)^{-\frac{1}{2}}d\sigma. \tag{4}$$

$F(p)$ is then determined as

$$F(p)=(\text{sign of } p)H(p^2) \tag{5}$$

which can be evaluated in terms of Fresnel integrals.

[4] R. V. L. Hartley, Bell System Tech. J. **29**, 350 (1950).

[5] M. Born, *Atomic Physics* (Blackie and Sons Limited, London, 1947).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

The Concept of Force in Dislocation Theory

M. O. PEACH
University of Notre Dame, Notre Dame, Indiana
(Received May 21, 1951)

The concept of "force acting on a dislocation" (or other imperfection) is intrinsically different from the concept of force as used in newtonian mechanics, and must be defined. A definition is given which preserves the basic property (1) that the direction of the resultant force gives the direction in which the imperfection will actually move if the atomic constraints do not prevent motion. For imperfections in an infinite crystal, and assuming linear elasticity, two further properties are established: (2) the parallelogram law of adding forces is valid, and (3) Newton's law of equal and opposite action and reaction is valid. A critical review of previous work relating to forces between crystal imperfections is given.

PART I. INTRODUCTION

THE introduction of the idea of force into dislocation theory has proved very useful, but it has also resulted in a certain amount of confusion. This confusion has led to a lack of agreement in computed values of force in certain situations; to a divergence of usage of the term force; and, in at least one case, to a fundamental misconception as to the behavior of a dislocation in a stress field. The discrepancies in computed values of force have not been serious, inasmuch as the calculations were in every case order of magnitude calculations. However, dislocation theory has advanced to a stage where more precise calculations are now possible, and such discrepancies are likely to assume a more serious aspect henceforth. In any event the divergence of usage ought to be reconciled, and the concept of force should be placed upon a firm theoretical foundation and its nature made clear.

Originally an intuitive concept founded in our immediate kinesthetic experience, the idea of force was adopted and made precise by newtonian mechanics, then later introduced into the description of certain nonmechanical phenomena. In this latter extension of usage the word has a meaning intrinsically distinct from its meaning in mechanics. For example, consider an infinite crystal containing two positive edge-dislocations, A and B. We know that they exert a repelling force on each other, and in the absence of restraining influences, will move apart. Yet if we surround A (or B) by any closed surface the "free-body" so formed will be in equilibrium. (We assume the crystal to be at rest, and we disregard thermal motions of the atoms.) In fact, explicit use is made of the existence of equilibrium in carrying through the computation which ultimately leads to a numerical value for the unbalanced "force" acting between A and B. The paradox is easily explained. The "force" acting between A and B is intrinsically different from the forces which act on the free-body. Ordinary forces act on masses. A dislocation is not a mass, but rather a geometrical configuration.

The possibility of extending the idea of force to nonmechanical phenomena rests upon two facts. (1) In all physical phenomena, mechanical or nonmechanical, energy changes occur. (2) In mechanical phenomena force can be defined in terms of energy changes.

PART II. FORCE AS A DERIVED CONCEPT IN MECHANICS

II(A). Example of a Mechanical System

A linear spring AC is fastened (Fig. 1) to a rigid support at A and to a cross at D. Three linear springs $B'D$, $A'D$ and $C'D$ are fastened to the cross at D and carry couplings at their ends to which the nonlinear springs BB' and CC' and the linear spring $A''A'$ can be fastened. Masses A'', B, and C are fastened to springs $A''A'$, BB', and CC', and rest upon a rough horizontal plane so that if they are placed in arbitrary (within limits) positions as shown in Fig. 2 they will remain fixed.

The spring system in Fig. 1 is a medium capable of storing potential energy and releasing it under appropriate conditions. In Fig. 2 the medium does possess potential energy, or field energy, and the forces exerted by the spring system on the masses A'', B, and C may be called field forces.

II(B). Definition of Field Force in Terms of Energy

Remove the rough constraint at C and replace it by a workless constraint permitting motion only in the given direction, together with an external force sufficient

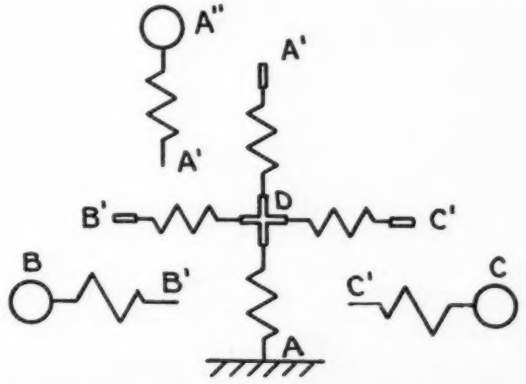


FIG. 1. Mechanical model of a perfect crystal $AB'A'C$; of an external set of tractions AA'; and of two imperfections BB' and CC'.

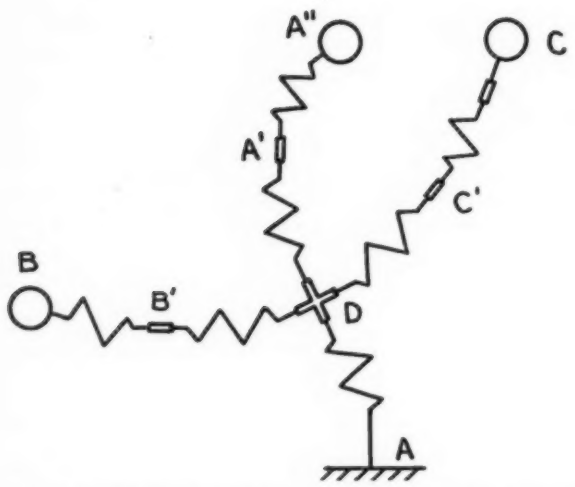


FIG. 2. Model of a crystal with two imperfections and external tractions. The system is fixed at A, and motion of A'', B, and C is prevented by the rough plane upon which they lie.

to maintain equilibrium. By reducing or increasing the external force an infinitesimal amount, permit a small displacement ds of C in the required direction. If the decrease of field energy which occurs in this displacement is $-dW$ then the component of field force exerted on C in the required direction is defined to be $-dW/ds$.

We now emphasize an important condition which must be satisfied if the definitions just given are to lead to a correct result for a mechanical system. The system must be closed everywhere except at the point where the force is being defined. In investigating the force on C we must make certain that in the small displacement the system loses energy only at C and not at A, B, or A''.

We note also that a correct result for the force would not be obtained if one considered only the energy stored in $AB'A'C'D$ and neglected the energy stored in the springs $B'B$, $A'A''$, and $C'C$. For, in the displacement of C, small displacements of B', A', and C' also occur, and energy is lost from the restricted system at these points.

II(C). Definitions of Forces of Constraint and of Dissipation in Terms of Field Force

First let us distinguish between statical and quasi-statical processes. A statical process is an equilibrium situation in which no motion occurs. A quasi-statical process is one in which the velocity changes are so small that kinetic energy changes may be neglected.

Then constraining force and dissipative force may be defined by the following statements.

In a statical process the forces of field and of constraint form an equilibrium system.

In a quasi-statical process the force of dissipation is oppositely equal to the component of field force in the direction of motion, and the forces of field, dissipation, and constraint form an equilibrium system.

II(D). Force Exerted by a Field-Producing Entity

The field-force defined in II(B) is the total force, in a given direction, exerted on the mass C by the total field. It was defined in terms of changes of total field energy. The field energy exists only by reason of the presence of the masses A'', B, C in their given positions, and, of course, the presence of the medium $ABA'C$ itself. If A'' and B were removed (e.g., by making the plane smooth in the neighborhood of A'' and B) there would still be energy in the medium and there would still be a force exerted upon C. This energy is usually called the self-energy of C and the force a self-force. (It might be more appropriate to refer to this energy as an interaction energy between C and the medium, and to the force as the force exerted on C by the medium.)

If the self-energy of C is subtracted from the total field energy the remainder is an interaction energy between C and the other sources of field energy A'' and B, and the force derived from this energy is the force exerted on C by A'' and B jointly.

PART III. GENERALIZED CONCEPT OF FORCE

III(A). Force on an Entity

If (a) there exists a field, i.e., a distribution of potential energy in a medium, and (b) if there exists an entity possessing position, and (c) if changes of field energy occur as the position of the entity is changed, then it is possible to define a quantity which can be interrupted as the force exerted on the entity by the field.

If there exists some mechanism which prevents free motion of the entity under the influence of the field force, then it is possible to define a force of constraint.

If quasi-static motion of the entity is possible then it is possible to define a dissipative force.

Whenever field force can be defined, then forces of interaction between field-producing entities can be defined.

One need only paraphrase Part II in an obvious manner to make it applicable to the general situation.

A crystal containing imperfections provides an example of a situation where the conditions of this section are satisfied.

III(B). The Parallelogram Law and Newton's Third Law

Consider first a closed system containing field-producing entities (1) at ξ_1, η_1, ζ_1, and (2) at ξ_2, η_2, ζ_2. We assume that the system is in a unique state and possesses a definite amount of stored configurational energy. The system can be brought into the above state, starting from an energy-free state in either of two ways; (2) can be created last, or (1) can be created last. Let W_1 be that portion of the energy used in creating (1) which is stored as configurational energy. Let $W_{1,2}$ be that part of the work done in creating (2)

in the presence of (1) which is stored as configurational energy. With similar definitions of W_2 and $W_{2,1}$, and using the basic assumption above, one can write for the total energy W of the system: $W_1+W_{1,2}=W_2+W_{2,1}$. Subtract W_1+W_2 from both members, obtaining $W_{1,2}-W_2=W_{2,1}-W_1$. Denote either of these equal quantities by W_{12}, where the order of subscripts is unimportant. Then $W=W_1+W_2+W_{12}$. It is clear that W_1 is independent of the position coordinates of (2) and W_2 is independent of the position coordinates of (1). In general W_{12} may contain position coordinates of both.

Repeat the above process with (2) in the same position but with (1) displaced slightly in the x-direction a distance $\xi_1'-\xi_1$. Using primes to denote the changed quantities one obtains $W'=W_1'+W_2'+W_{12}'$ and then

$$-\frac{W'-W}{\xi_1'-\xi_1}=-\frac{W_1'-W_1}{\xi_1'-\xi_1}-\frac{W_{12}'-W_{12}}{\xi_1'-\xi_1}.$$

The quantities appearing (more precisely, their limits as ξ_1' approaches ξ_1) are, respectively, the x components of the total force on (1), the self-force on (1), and the force exerted on (1) by (2). Repeat again, holding (1) fixed in its original position and giving (2) a slight displacement $\xi_2''-\xi_2$ taken equal to $-(\xi_1'-\xi_1)$.

Then, using double primes to indicate the new quantities,

$$+\frac{W''-W}{\xi_1'-\xi_1}=-\frac{W_2''-W_2}{\xi_2''-\xi_2}-\frac{W_{12}''-W_{12}}{\xi_2''-\xi_2}.$$

Comparing the last two equations one sees that if the self-forces are negligibly small and if $(W''-W)/(\xi_1'-\xi_1)$ is negligibly small then Newton's third law is valid. But this latter quantity is the self-force exerted on the group of two entities regarded as a single composite entity. For in the displacement from state W' to state W'' the two entities move together in a rigid-body displacement. Accordingly we may say that a sufficient condition for the validity of Newton's third law is that the self-forces of the entities, both individually and as a group, be negligible. The question of the validity of the parallelogram law does not arise since only two entities are being considered.

Consider now three field-producing entities. Proceeding as above and with an obvious extension of the notation one obtains $W=W_1+W_2+W_3+W_{12}+W_{23}+W_{31}+W_{123}$ where W_{12}, for example, is identical to the W_{12} used above, while W_{123} is a new quantity which may contain the coordinates of all three entities. The mutual forces between (1) and (2) arise from W_{12} whereas the force on (1) arising from W_{123} is a joint force involving also (3). *Hence the presence of (3) has no effect on the validity of Newton's third law as between (1) and (2).* It can be seen that additional entities would not alter this result.

Now give (1) a small displacement, holding (2) and (3) fixed. Then: $-\partial W/\partial\xi=-\partial W_1/\partial\xi-\partial W_{12}/\partial\xi-\partial W_{31}/\partial\xi-\partial W_{123}/\partial\xi$. Transpose the self-force term; the equation then states that the force exerted on (1) by (2) and (3) together is the sum of the forces which they would exert separately plus an additional term. In other words: A necessary and sufficient condition that the parallelogram law of addition be valid is that such quantities as W_{123}, W_{1234}, $\cdots$, if they exist at all, be independent of the position coordinates of the entities.

PART IV. APPLICATION TO A CRYSTAL CONTAINING IMPERFECTIONS

IV(A). Mechanical Model of an Imperfect Crystal

Sufficient generality for our purpose will be obtained by considering a crystal containing an edge dislocation and a foreign atom and subjected to an external set of tractions at its surfaces (Fig. 3). Some part of the surface, say A, will be fixed or supported in some manner and the remainder A' will have a set of tractions on it. There will in most cases be a region of nonelastic stress surrounding each imperfection, such as the foreign atom at B and the edge dislocation C. Corresponding to this crystal we have the model shown in Fig. 1. The fixed portion of the external surface is at A and the remainder at A'. The nonlinear spring BB' and the mass B represent the foreign atom and surrounding region of nonelastic stress. Similarly $C'C$ and C represent the dislocation. The externally applied tractions are exerted by $A'A''$. The rough plane serves to represent the atomic constraints which oppose the motion of the imperfections B and C. This model should be kept in mind while reading Part V.

IV(B). Force Exerted on a Crystal Imperfection by a Stress Field

In Fig. 3 there will be a stress field τ in the crystal arising from A', B, and C. To find the force exerted on C by τ in the direction **ds** we must calculate the energy lost $-dW$ as C moves quasi-statically a distance ds to C_1. In this calculation the energy changes in the spring system must be included, since energy must be allowed

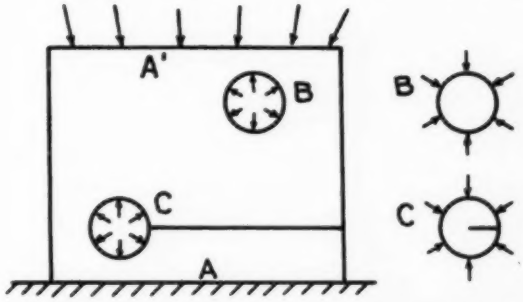


FIG. 3. Crystal with external set of tractions and two imperfections. The nonelastic regions are imagined to be removed as shown, with an outside agency exerting the tractions necessary for equilibrium.

to enter or leave the system only at C. To carry out the calculation note that every imperfection has associated with it a characteristic parameter λ which measures its "strength." For example, in the case of a dislocation this parameter is the magnitude of its Burgers vector; for a foreign atom, it is the fractional increase in radius of the foreign atom as compared to the radius of a normal atom. The energy change as the imperfection moves from C to C_1 may be computed as the sum of the energy changes which occur in the following process: Let the imperfection be reduced to zero strength by slowly decreasing λ to zero. Let the imperfection of zero strength be moved from C to C_1. Let the strength be gradually increased to its former value.

An example of such a calculation is given in a previous paper.[1]

IV(C). Forces Exerted by Crystal Imperfections on Each Other

It can be shown that Newton's third law is valid as between imperfections in an infinite crystal, i.e., in a "large" crystal where the imperfections are not too "near" the surface. Consider first a line-localized imperfection, such as an edge dislocation, in a large cylinder of radius R. Moving the imperfection a small distance $d\xi$ is equivalent to holding the imperfection fixed and displacing the cylindrical boundary a distance $-d\xi$. In this process energy is lost from a small region of crescent-shaped cross section on one side of the crystal and gained in a similar region on the opposite side. The net change in energy is certainly less than that stored between two cylindrical surfaces of radii R and $R+d\xi$. Since the stresses fall off inversely as the first power of the distance from the imperfection the latter energy goes as $2\pi R d\xi/R^2$. Hence the self-force is certainly exceeded by a quantity of order R^{-1} and for large R may be neglected.

For an imperfection not extending clear across the crystal but occupying a finite volume surrounding a point, the stresses go as R^{-3}. The energy stored between two spherical shells of radii R and $R+d\xi$ goes then as $4\pi R^2 d\xi/R^6$. Hence we may conclude that Newton's third law holds for imperfections in an infinite crystal.

The validity of the parallelogram law of adding forces depends upon the non-occurrence of terms of the type W_{123}, W_{1234}, $\cdots$ discussed in III(D). If there is a linear relation between stress and strain (or, more fundamentally, if there is a linear force-displacement law for the interaction of atoms) then such quantities will not appear. Since Hooke's law does not extend to large strains, and since, in general, there are always regions where the strains are large, we may say that, in a very strict sense, the parallelogram law is not valid. However, the regions where large strains occur are generally extremely small in comparison to the regions where elastic strain occurs, so that the terms W_{123}, etc., are negligibly small as a general rule. It appears that the parallelogram law is valid to a very good approximation for the imperfections which have been introduced into the theory up to the present time.

[1] M. O. Peach and J. S. Koehler, Phys. Rev. **80**, 436 (1950).

PART V. DISCUSSION OF THE LITERATURE RELATING TO FORCE

(A) Mott and Nabarro[2] (1948) gave the correct result that the force per unit length on an edge dislocation caused by an externally applied shear stress σ in the slip direction and on a plane parallel to the slip plane, is $\sigma\lambda$ where λ is the magnitude of the Burgers vector. The proof was essentially qualitative, depending upon the work done in moving the dislocation clear through the crystal. Strictly speaking, it was proved that the above result was the average force on the dislocation during the entire motion.

(B) T. W. Read and W. Shockley[3] (1950) defined the force on a dislocation as the energy dissipated in moving the dislocation through unit distance on the slip plane. On the basis of this definition they deduced the result of (A) above by a semi-intuitive procedure somewhat similar to that used by Mott and Nabarro and subject to the same qualification pointed out above. They then proceeded to give a rigorous proof of the result, and pointed out that it was valid not only for an edge dislocation but for any straight line dislocation.

The definition given has obvious limitations. If the dislocation is not moving the externally applied shear stress exerts no force on it. The force exerted by externally applied stresses other than shear is not given. In any case no force component exists except in the slip plane.

In the light of our discussion in Part III the true nature of this definition becomes clear. Read and Shockley have given not a definition of the force exerted on a dislocation by a field, but rather a definition of the force exerted on a dislocation by dissipative constraints. This becomes equivalent to the definition given by us for the special case in which the dislocation is moving quasi-statically and forces exerted by workless constraints are zero.

(C) Koehler[4] (1941) defined force as the negative gradient of the strain energy stored in the crystal. Since the stress fields used for a dislocation required tractions at the external surface, this must be regarded as incorrect on the grounds that the elastic region wherein the strain energy was stored did not constitute a complete system. The result deduced for the force on a dislocation caused by an externally applied shear stress and that for the force between two edge dislocations differ somewhat from the correct results. None of the major conclusions of the paper are affected by this discrepancy.

[2] N. F. Mott and F. R. N. Nabarro, Phys. Soc. (London), Bristol Conference Report (1948), p. 3.
[3] W. Shockley and R. W. Read, Phys. Rev. **78**, 275 (1950).
[4] J. S. Koehler, Phys. Rev. **60**, 397 (1941).

(D) C. Leibfried[5] (1949) considered separately the force exerted on an edge dislocation by an externally applied stress, and that exerted by a stress field caused by some other imperfection within the crystal.

(1) The force on a dislocation caused by external means is defined as the negative gradient of the strain energy subject to the condition that the points of the external surface where forces are applied are to be held fixed. The effect of the constraining condition is to insure that no work is done at the surface of the crystal, i.e., to close the system at the surface. This definition is essentially equivalent to ours except that we do not insist that the system be closed in a particular manner; any way will do.

The results obtained for an edge dislocation and an arbitrary external stress system are shown to apply also for a screw dislocation.

(2) In the case of the force exerted on a dislocation by some other internal imperfection Leibfried apparently reverts to Koehler's 1941 definition of force, i.e., he takes the negative gradient of the interaction strain energy stored in the crystal. For the case of two dislocations he arrives at the same expression for the force in terms of the stress field as above (*D*)(1). The reason is clear: the system is closed. For the case of an edge dislocation and a foreign atom he concludes that the behavior is essentially different; that the stress field produced by a foreign atom affects the dislocation in a different way than that produced by another dislocation or an externally applied stress system.

Leibfried obtains

$$\phi = -4G\lambda R^2(R-r)\eta/(\eta^2+\xi^2), \tag{1}$$

by integrating over the cut surface S (see Fig. 4) the product of the displacements caused by the dislocation and the tractions caused by the stress field of the foreign atom. (Here G is the shear modulus, λ the magnitude of the Burgers vector, r and R are the radii of normal atom and foreign atom, respectively.) Then this product is integrated over the internal surface I of the crystal located at the foreign atom to obtain

$$\phi' = \frac{8(1-2\sigma)}{3(1-\sigma)} G\lambda R^2(R-r)\frac{\eta}{\eta^2+\xi^2}. \tag{2}$$

(Here σ is Poisson's ratio.) Then (2) is added to (1) to obtain the total interaction energy. Our contention is that (2) should not be taken into account since in order to have a closed system the foreign atom itself must be included, and then the transfer of the energy (2) from one part of the system to another is a matter of no consequence.

Leibfried states that the foreign atom is thought of as incompressible, but he does not treat it so in his calculation. Had he done so, his treatment of the foreign atom would have exactly paralleled his treatment of an external set of tractions on the surface of the crystal, and he would have been led to (1) for the total interaction energy, a result in complete agreement with his previous results regarding the influence of a stress field on a dislocation.

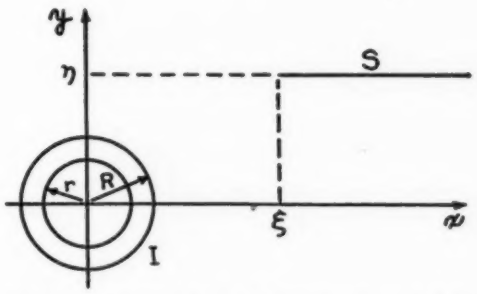


FIG. 4. Foreign atom and edge dislocation in an infinite crystal (after Leibfried). S is the cut-plane of the dislocation. The foreign atom has a radius R, different from the radius r of a normal atom.

(E) Cottrell[6] (1947) obtained for the interaction energy of a foreign atom and an edge dislocation the result,

$$\phi'' = -\frac{4(1+\sigma)}{3(1-\sigma)} G\lambda R^2(R-r)\frac{\eta}{\eta^2+\xi^2} \tag{3}$$

by integrating over the internal surface I of the crystal (Fig. 4) the product of the displacements caused by the foreign atom and the tractions caused by the stress field of the dislocation. This result neglects the interaction energy stored in the foreign atom, which is the negative of (2) above. If this is taken into account the total interaction energy becomes (1) above, as indeed it must, since we know Newton's third law is valid for the present case. Bilby[7] has verified this by direct calculation. None of the major conclusions of the paper are affected by this discrepancy.

(F) Bilby[7] (1950) has given the following definition: "The interaction energy of a given source in a stress field is the change in energy of the system, including both work done by the external forces and changes in the elastic energy of the crystal, which accompanies the introduction of the source." If by system is meant the elastic regions of the crystal, and by external forces is meant the forces exerted by the system on objects not included in the system, then the definition leads to a definition of force which is equivalent to that proposed in the present paper. Since the words "system" and "external forces" appear only in the above statement and occur nowhere else in the paper, it is not clear to the present writer exactly what was intended.

Bilby reviews results previously obtained for the interaction of two edge dislocations, and an edge dislocation and foreign atom, noting that in these calcula-

[5] G. Leibfried, Z. Physik **126**, 761 (1949).

[6] A. H. Cottrell, Report of a Conference on the Strength of Solids, The Physical Society, London (1948).

[7] B. A. Bilby, Proc. Phys. Soc. (London) Sec. A, **63**, 3 (1950).

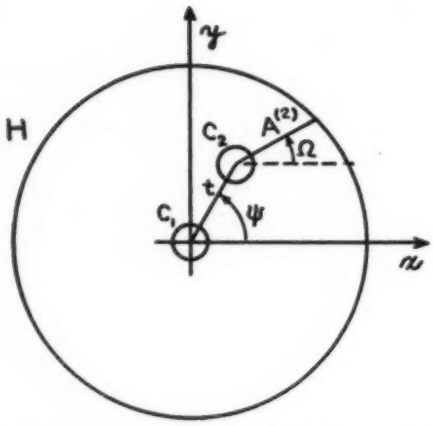


FIG. 5. Two edge dislocations in an infinite cylinder (after Bilby). A is the cut-plane of dislocation 2.

tions the stresses for an edge dislocation in an infinite medium with tractions applied to the inner boundary were used. He then proposes an investigation in which the stress formulas for the edge dislocation are modified to require zero tractions on the external surface H of the crystal and on the internal surface C surrounding the dislocation line. The two edge dislocations are taken in an infinite cylinder (Fig. 5), one at the origin and the second at t, ψ. The cut surface $A^{(2)}$ for the second one makes an angle Ω with the x axis. The Burgers vectors have components $(\lambda_1 l_1, 0, 0)$ and $(\lambda_2 m_1, \lambda_2 m_2, 0)$ respectively. The stresses for dislocation (1) are obtained from the stress function,

$$\chi^{(1)}=\gamma\left[-R\ln R+\frac{R^3}{2(L^2+r_0^2)}-\frac{L^2r_0^2}{2(L^2+r_0^2)}\frac{1}{R}\right](\lambda_1 l_1\sin\Phi),$$

where (R, Φ) locate the field point. The three terms in this expression are, respectively, (a) the one used by previous investigators, which gives the stresses of an edge dislocation in an infinite medium, (b) the term which makes the tractions zero on the external boundary H and, (c) the term that makes the tractions zero on the internal boundary C_1. Bilby then computes the interaction energy by integrating the product of the stresses caused by dislocation (1) and the displacements caused by dislocation (2) over the surfaces $A^{(2)}$, H, C_1, and C_2.

According to our conceptions, this integration should have been taken over $A^{(2)}$ only.

The radius R of the crystal is allowed to become infinite and it is shown that the integrals over H, C_1, and C_2 become negligible.

Our contention is that even if these integrals had not turned out to be negligible the correct answer would still be given by the integral over $A^{(2)}$ alone.

In the integration over $A^{(2)}$ it is found that the (c) term is negligible; the (b) term gives

$$W_2=\lambda_1\lambda_2 l_1\gamma[m_1(-\tfrac{1}{2}-\sin^2\Omega)+m_2(\tfrac{1}{2}\sin 2\Phi)], \quad (4)$$

and the (a) term gives

$$W_1=\lambda_1\lambda_2 l_1\gamma[m_1(\ln L-\ln t-\sin^2\psi+\sin^2\Omega) +m_2(\tfrac{1}{2}\sin 2\Phi-\tfrac{1}{2}\sin 2\Omega)]. \quad (5)$$

It is stated that (5) gives the correct force expressions in all cases except when $\Omega=\psi$, in which case (4) must also be included. This is apparently a mathematical error arising from setting the constant Ω equal to the quantity ψ which, in the differentiation to obtain the force, is assumed to be a variable. Since (4) does not contain the position coordinate of the dislocation it is clear that it can not make a contribution to the force under any circumstances.

The treatment of two screw dislocations proceeds similarly to that above. It is shown that the "accurate" expression reduces to an integral over the cut surface of one of the dislocations. Our contention is that if it did not do so, then the "accurate" expression would be wrong and the cut integral would be correct.

The treatment of an edge dislocation and a foreign atom proceeds similarly to those above until Cottrell's result (3) is obtained. Then the energy stored in the foreign atom is computed, which turns out (as it must) to be the negative of an integral already computed over the surface of the hole containing the foreign atom. When this energy is added (that is, when the previously computed integral is struck off) the correct result (1) is obtained.

(G) Peach and Koehler[1] (1950) derived a general expression for the force on a general Burgers dislocation in any stress field whatever. No formal definition of the force on a dislocation was given but the procedure used may be taken to imply a definition equivalent to the one laid down in the present paper. All the conclusions of that paper are correct on the basis of this definition.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

Atomic Clocks and Frequency Stabilization on Microwave Spectral Lines*

C. H. Townes†
Columbia University, New York, New York
(Received May 21, 1951)

Application of the various types of radiofrequency spectral lines to accurate frequency stabilization and time standards is surveyed. Pertinent characteristics of microwave gas absorption lines and the various types of errors in frequency stabilization due to the nature of these absorption lines or to fundamental thermal noise are discussed in detail. It is shown that time standards synchronized with microwave absorption in ammonia or resonances in molecular or atomic beams have limits of accuracy of the order of 1 part in 10^{12} for a short time, and still smaller limiting fractional errors over longer periods of time.

INTRODUCTION

FOR a long time physicists have been aware of the attractiveness of an "atomic clock," i.e., a clock controlled by some periodic motion in the atomic world. Periodic motion of macroscopic matter, such as are traditionally used for measuring time intervals, are all subject to variations because macroscopic matter is complex and subject to many varieties of small changes. Even the period of rotation of the earth is known to have varied as much as three parts in 10^8 during the last 50 years. Since atomic or molecular systems are made of a small number of fundamental particles, their periodic motions tend not to be subject to small changes. The spectral lines of an ammonia molecule or a cesium atom, for example, might be expected to remain unchanged unless something rather violent occurs such as a disruption of the molecule or atom. Only in recent years have radio techniques and known sharp molecular and atomic lines overlapped in frequency so that electrical devices and therefore mechanical clocks can be controlled by spectral line frequencies.[1]

The known sources of sharp radiofrequency lines are:

1. Nuclear precession in an external magnetic field.
2. Nuclear quadrupole coupling to electrostatic fields in solids.
3. Molecular rotational and other frequencies detected by absorption of microwaves.
4. Hyperfine structure and other resonances detected by atomic or molecular beam techniques.

Resonances of type 1 have frequencies which are proportional to the externally applied magnetic field and hence may be variable. Resonances of type 2 may give useful frequency standards in the range of a few hundred megacycles, but their characteristics are not yet very completely known.‡

Both type 3 and 4 resonances are very promising for frequency standard and clock applications, and use of resonances of type 3 has already been discussed.[1] It is our purpose to analyze some of the various characteristics of the resonances themselves which affect the performance of electronic oscillators synchronized to these resonances. By far the most attention will be given to microwave absorption resonances (3) because they are the most popular type for frequency stabilization and are much more susceptible to a general treatment.

WIDTH OF SPECTRAL LINES

A truly isolated, undisturbed, and stationary molecular system would have the attractive feature of definite and fixed energy levels, but various types of unavoidable disturbances do in fact vary the energy levels, giving a width to spectral lines and varying their average or center frequencies. The sources of spectral line broadening which need to be considered are:

1. Natural line breadth.
2. Doppler effect.
3. Pressure broadening, i.e., disturbances due to collisions between molecules.
4. Collisions between molecules and the walls of a containing vessel.
5. Saturation broadening.

The natural line breadth may be interpreted classically as due to radiation damping, or quantum mechanically as a disturbance of the molecule by zero-point vibration of electromagnetic fields which are always present in free space. For a transition of frequency ν from an excited state to the ground state of a system, zero-point electromagnetic fields give an absorption line a half-width at half-maximum intensity of

$$\Delta\nu = (32\pi^3\nu^3/3hc^3)|\mu|^2, \quad (1)$$

where μ is the quantum-mechanical matrix element of the dipole moment-usually of the order 10^{-18} esu. For radiation of 1 cm wavelength, $\Delta\nu$ is from (1) approxi-

* Work done under the auspices of the National Bureau of Standards.
† Consultant Physicist, National Bureau of Standards.
[1] E. W. Fletcher and S. P. Cooke, Cruft Laboratory, Harvard University Technical Report No. 64, (May 15, 1950); W. D. Hershberger and L. E. Norton, RCA Rev. 9, 38 (1948); H. Lyons, Phys. Rev. 74, 1203 (1948)T; 76, 161 (1949); Har. Instit. Amer. Jour. 4, 11, Feb and 4, 7, March (1949); H. Lyons, J. Appl. Phys. 21, 59 (1950); J. L. G. De Quevedo and W. V. Smith, J. Appl. Phys. 19, 831 (1948); Smith, De Quevedo, Carter, and Bennett, J. Appl. Phys. 18, 1112 (1948); Townes, Holden, and Merritt, Phys. Rev. 74, 1113 (1948).

‡ For a discussion of some of their predictable properties, see R. V. Pound, Phys. Rev. 79, 685 (1950).

mately 10^{-7} cycle/sec. For radiofrequencies and ordinary temperatures, thermal radiation consists of stronger electromagnetic fields than the usual zero-point fields, since each mode of vibration of the field has a mean energy kT rather than $\frac{1}{2}h\nu$. Allowing for broadening of both ground and excited states, this increases the value of $\Delta\nu$ by a factor $4kT/h\nu$ or approximately 800 for room temperature, giving a "natural" width of 10^{-4} cycle/sec. Although natural line breadths are of considerable importance in spectral regions of higher frequency because of their dependence on ν^3, this breadth of 10^{-4} cycle/sec is completely negligible in comparison with other broadening processes to be discussed.

The Doppler effect occurs when a molecule is moving parallel to the direction of propagation of the radiation being absorbed, and gives a frequency shift of $\pm\nu(v/v_p)$ where ν is the resonant frequency without Doppler shift, v the molecular velocity, and v_p the velocity of phase propagation of the radiation. Although under some conditions v_p may be larger than c, e.g., for propagation in a wave guide near cutoff, usually $v_p \approx c$, and the fractional frequency shifts are simply v/c. The probability that a molecule in a gas at temperature T has a velocity v in a particular direction is proportional to $\exp(-mv^2/2kT)$, where m is the molecular mass. Hence the line intensity as a function of change ϵ from the resonance frequency is $\exp[(-mc^2/2kT)(\epsilon/\nu)^2]$. The line is consequently symmetric and has a half-width at half-maximum of $\Delta\nu=(\nu/c)((2kT/m)\log_e 2)^{\frac{1}{2}}$. For an ammonia molecule at room temperature, $\Delta\nu/\nu = 1.5\times10^{-6}$. Doppler effect can be decreased to some extent by use of heavier molecules and lower temperatures, but a decrease in line width of more than a factor of two can hardly be expected because at low temperatures molecules do not have sufficient vapor pressure ($\sim10^{-2}$ mm of Hg) to absorb radiation.§

Although the Doppler effect gives the same fractional breadth $\Delta\nu/\nu$ for any frequency, the effects of intermolecular collisions give a fixed breadth $\Delta\nu$ independent of the resonant frequency and hence are considerably more noticeable in the microwave region than in other spectral regions. The line shape and intensity are given to a good approximation by Van Vleck and Weisskopf's expression.[2]

$$\gamma=\frac{8\pi^2Nf}{3ckT}|\mu|^2\nu^2\left[\frac{1/2\pi\tau}{(\nu-\nu_0)^2+(1/2\pi\tau)^2}+\frac{1/2\pi\tau}{(\nu+\nu_0)^2+(1/2\pi\tau)^2}\right] \quad (2)$$

where,

$1-e^{-\gamma}\approx\gamma$ is the fraction of power of frequency ν absorbed per centimeter from a plane wave,
N is the number of molecules per cc,
f is the fraction of molecules in the lower state of the transition of resonant frequency ν_0,
$|\mu|^2=|\mu_x|^2+|\mu_y|^2+|\mu_z|^2$ is the square of the dipole matrix element for the transition,
τ is the average time between collisions for an individual molecule.

At pressures less than 1 mm of Hg the second term in the brackets of (2) is not of great importance, and the line shape due to collision effects becomes that of a typical resonance curve,

$$\gamma=(8\pi^2Nf/3ckT)|\mu|^2\nu^2\Delta\nu/[(\nu-\nu_0)^2+(\Delta\nu)^2]. \quad (3)$$

The line half-width $\Delta\nu$, which is proportional to the number of collisions per second, is, of course, directly proportional to the gas pressure if the pressure is low and if the temperature is held fixed. Unfortunately, its value, dependence on molecular parameters and on temperature cannot be accurately predicted by theory. Nor can it be deduced from any experimental measurement other than a direct measurement at some pressure for the absorption line in question. A very rough value of $\Delta\nu$ for many gases is 15 megacycles/sec for a gas pressure of 1 mm of Hg.

An important result of (3) is that as long as pressure broadening is the primary source of line breadth, the value of γ at resonance ($\nu=\nu_0$) is independent of pressure. For then $\gamma=8\pi^2Nf\nu^2/3ckT\Delta\nu$, and both N and $\Delta\nu$ are proportional to pressure.

Absorption of radiation by a molecule may be disturbed either by a collision with another molecule, or with the walls of a containing vessel. The average number of collisions per second per molecule with the walls is given by $A/V(RT/2\pi M)^{\frac{1}{2}}$ where A is the wall area, V the volume of the inclosure, M the molecular weight of the gas, and R the gas constant. An absorption line would have a half-width due to these collisions of approximately $\Delta\nu=(A/2\pi V)(RT/2\pi M)^{\frac{1}{2}}$. For ammonia gas at room temperature in a long K-band wave guide (cross section 1 cm by 0.5 cm), the resulting half-width is 7 kilocycles. Although this source of line broadening is not negligible, its effect is usually small compared with other types of broadening.

Still another cause of line broadening, and one peculiar to radiofrequency spectroscopy, is saturation[3] of

§ G. Newell and R. H. Dicke, Phys. Rev. **81**, 297(A) (1951), have developed a technique not yet described in detail, of selecting only molecules in a certain narrow velocity range, and thus decreasing the effect of Doppler broadening on the line width by a factor of ten or more. This technique may prove very useful, but it does not improve the theoretical limit of accuracy of stabilization. This is because the pressure must be lowered so that pressure broadening is considerably less than Doppler broadening, and in this process the absorption intensity is decreased. In addition, the power traversing the gas must be decreased to avoid saturation, which further decreases the intensity of the observed signal. The net result is a narrower but weaker line with an actual decrease in theoretical accuracy of stabilization. In some practical cases, however, their technique may have advantages.

[2] J. H. Van Vleck and V. F. Weisskopf, Revs. Modern Phys. **17**, 227 (1945).

[3] C. H. Townes, Phys. Rev. **70**, 665 (1946); B. Bleaney and R. P. Penrose, Proc. Phys. Soc. (London) **60**, 83 (1948); R. Karplus and J. Schwinger, Phys. Rev. **73**, 1020 (1948); H. S. Snyder and

the ability of the gas molecules to absorb due to large radiation intensities. This saturation is not connected with a distortion of the molecule because of excessive excitation by the electromagnetic wave as might be expected from classical theory. Each molecule which is in a state to absorb the particular frequency of interest does so only by being promoted to a higher energy state in which it can absorb no more energy of this frequency. After an average time t, called the relaxation time, this molecule may make a collision and return to its original state. If the rate of absorption of quanta is comparable with or greater than the relaxation rate, then the gas absorption coefficient is no longer independent of radiation intensity, but decreases as the intensity is increased. As pointed out above, as long as pressure broadening dominates, the absorption coefficient at resonance is independent of pressure. This means that each molecule absorbs more quanta per second as the pressure is decreased. Furthermore, since redistribution of energy after absorption depends on collisions, the relaxation time increases as pressure is decreased. Hence the power at which saturation effects become evident is proportional to the square of the pressure. For a typical case, saturation may be observed at a gas pressure of 10^{-2} mm of Hg with a radiation flux of 1 milliwatt per square centimeter.

A general expression for the absorption coefficient, replacing (3) if power saturation occurs, is

$$\gamma=\sum_M \frac{8\pi^2 Nf}{3ckT} \times \frac{|\mu_M|^2\nu^2\Delta\nu}{(\nu-\nu_0)^2+(\Delta\nu)^2+(8\pi^2/3ch)|\mu_M|^2\nu\cdot I\cdot t\cdot\Delta\nu}, \quad (4)$$

where the summation is over the various magnetic quantum numbers M or molecular orientations and $|\mu_M|^2$ is the square of the dipole matrix element $|\mu_z|^2$ for a particular value of M. This summation is necessary because molecules of different orientations saturate at different values of radiation intensity I. However, the sum may be approximately represented by one term with a typical value of the dipole matrix element $|\mu_M|$ in the denominator and $\sum|\mu_M|^2=|\mu|^2$ in the numerator. Here t is the relaxation time and I is the radiation intensity in terms of number of quanta per second per unit cross-sectional area. It has been found experimentally[3,4] that at least in the case of ammonia, each molecular collision which is effective in broadening the microwave absorption lines is also effective in redistributing the energy absorbed by the molecule, so that $t=1/2\pi\Delta\nu$. If the radiation frequency is varied slowly over the absorption line, and if we approximate (4) by using only one term in the summation, the line shape from (4) is again a resonance curve as in (3), but with a half half-width of

$$\Delta\nu(1+(8\pi^2|\mu_M|^2\nu It/3ch\Delta\nu))^{\frac{1}{2}}. \quad (5)$$

Here $|\mu_M|^2$ represents a typical value of $|\mu_M|^2$. Thus saturation is noticeable if $(8\pi^2|\mu_M|^2\nu It/3ch\Delta\nu)$ becomes as large as unity or $I\gtrsim(3ch(\Delta\nu)^2/4\pi|\mu|^2\nu)$. Here t has been set equal to $1/2\pi\Delta\nu$, which is actually only its upper limit, and $|\mu_M|^2$ has been approximated by $|\mu|^2$.

VARIATION OF APPARENT SPECTRAL FREQUENCIES

Having discussed the various causes of damping or of line width, we can now turn to an examination of the apparent constancy of line frequency. As long as a molecule is completely free and stationary, its frequencies are immutable—in pleasant contrast to frequencies of larger systems. Actually, however, the molecule must be contained as a gas and will be subject to some variations of pressure and temperature as well as to the influence of miscellaneous external electric, magnetic, and gravitational fields.

Expression (2) allows no change in the theoretical resonance frequency ν_0 with pressure or temperature. However, any device stabilized on the frequency of the absorption line would hardly be able to take into account the entire shape of the line and deduce ν_0. Rather, such a device would probably choose as the fiducial frequency either the frequency where maximum absorption occurs, or the average of two frequencies on the steep slope of the absorption curve which show identical absorption. These frequencies will be slightly affected by the unsymmetrical shape of the absorption curve and hence by changes in $\Delta\nu$ or $1/2\pi\tau$, due to variation in pressure or temperature.

The frequency for maximum absorption when $\Delta\nu\ll\nu_0$ is given by

$$\nu=\nu_0[1+(\Delta\nu/\nu_0)^2]. \quad (6)$$

For a typical case, $\Delta\nu/\nu_0$ might be 1/50,000 and a 10 percent change in $\Delta\nu$ due to a 10 percent change in pressure would correspond to a variation in the frequency ν of 8 parts in 10^{11}. For lower pressures or smaller $\Delta\nu$, this variation would be smaller. Temperature variations would be less troublesome than pressure variations. A 10°C variation in temperature at room temperature would not vary $\Delta\nu$ more than 2 percent and hence the apparent frequency would be constant to 1.6 parts in 10^{11}.

If the resonant frequency is taken as the average of two frequencies on the steepest slope of the absorption line which undergo the same absorption, then the frequency obtained is

$$\nu=\nu_0[1+\tfrac{1}{2}(\Delta\nu/\nu_0)^2]. \quad (7)$$

For this case, a 10 percent change in pressure gives half as much shift as for the previous case, or 4 parts in 10^{11}.

There is, however, a somewhat more insidious type

[3] P. I. Richards, Phys. Rev. **73**, 1178 (1948); R. L. Carter and W. V. Smith, Phys. Rev. **73**, 1053 (1948).
[4] R. Karplus, Phys. Rev. **74**, 223 (1948).

of frequency variation which accompanies variation with pressure. The Van Vleck-Weisskopf expression (2) for the shape of an absorption line is strictly correct only under the assumptions that

a. Collisions between molecules take place in a time very short compared with the period of oscillation of the radiation.

b. Each collision is violent enough so that after the collision the phase of the molecular oscillation is completely independent of its phase immediately before collision.

For frequencies in the infrared region, it is well known that these assumptions are not correct, and the equivalent of expression (3) becomes

$$\gamma = 8\pi^2 N f/3ckT|\mu|^2\nu^2\Delta\nu/(\nu-\nu_0-\alpha\Delta\nu)^2+(\Delta\nu)^2 \quad (8)$$

where α is of the order of 1 or $\frac{1}{10}$. Thus the apparent resonant frequency shifts due to intermolecular forces by an amount that is proportional to pressure and is of the order of $\Delta\nu$. For all lines studied in the microwave region, the constant α has been undetectably small—certainly less than 1/20 at low pressures.[5] However, even for a value of α as small as 1/20, a 10 percent change in pressure would correspond to a frequency change of 1 part in 10^7 if $\Delta\nu/\nu_0$ is 1/50,000. It is, therefore important to study the magnitudes of the constant α for various cases if high stabilization accuracy is to be achieved. Probably it will be possible to choose an absorption line for which α is extremely small, but better methods of measuring this shift in resonant frequency must be devised before a good choice of an absorption line for a high accuracy frequency standard can be made.

For ammonia, the only shift in resonant frequency which has so far been detected is of a somewhat different type. It occurs importantly only at high pressure, and is proportional to the square of the pressure. Margenau[6] has discussed the theory of this effect and given the formula $\nu=\nu_0(1-0.36p^2)$ for the apparent resonant frequency, where p is the pressure in atmospheres. Therefore, for 10^{-2} mm of Hg pressure, a 10 percent change in pressure would produce a shift of 3 parts in 10^{11}.

Stark effects, i.e., shifts in resonant frequency with application of an electric field, are very easily obtained in the microwave region. Certain special types of absorption lines, such as those produced by symmetric top molecules, are extremely sensitive to electric fields and split symmetrically into a number of components separated by a number of megacycles with the application of fields as weak as a few volts per centimeter. More characteristically, microwave absorption lines split up unsymmetrically with the application of an electric field, the magnitude of the splitting being proportional to E^2, the square of the field strength, and requiring several hundred volts per centimeter for a splitting of a few megacycles. In the case of the strongest ammonia line (the 3, 3 line at 23,870 Mc) the absorption is split into several component lines,[7] the strongest of which is at a frequency in megacycles $\nu=23{,}870+2.5\times10^{-5}E^2$. Thus, for $E=200$ volts/cm, its frequency is 23,871 Mc. For a displacement due to E less than one part in 10^{12}, E must be kept less than 1/30 volt/cm. Since normally microwave absorption would occur inside a conducting wave guide, frequency shifts due to spurious electric fields present no serious problem.

Fortunately, stray magnetic fields, such as the earth's field, produce essentially no shift in the frequency of microwave lines for normal molecules. Most molecules are in $^1\Sigma$ states for which the electronic magnetic moments completely cancel. The molecules therefore have only very small magnetic moments of the order of a nuclear magneton. In addition, the rather small magnetic effects which are present are almost completely symmetric, so that they only broaden the line slightly without shifting its center. The amount of broadening for $^1\Sigma$ type molecules is roughly $\mu_n H/h$; where μ_n is a nuclear magneton, H the magnetic field in oersteds, and h is Planck's constant. Numerically, this is approximately 1000 H cycles/sec, and hence is negligible. However for certain unusual molecules such as O_2 which are not in $^1\Sigma$ states, the splitting or broadening is roughly $\mu_e H/h$, where μ_e is a Bohr magneton, which is 1837 times larger than μ_n. In these cases, even the earth's magnetic field of approximately 1 oersted can produce a splitting of nearly 1 Mc.

A uniform gravitational field does not effect the apparent frequencies of molecular or atomic systems if the frequency measurement is made at the same gravitational potential as the oscillating system. A non-uniform gravitational field could in principle change molecular frequencies, but this effect is actually insignificant. The most interesting and significant effect of gravitational fields on frequencies is connected with the principle of general relativity that if two identical clocks (or oscillating systems) at gravitational potentials which differ by $\Delta\phi$ are compared, the one at lower gravitational potential appears to run more slowly by a fractional amount $\Delta\phi/c^2$. This change may be explained on a rather simple basis.[8] If time signals in the form of photons of frequency ν, energy $h\nu$, and mass $h\nu/c^2$ are sent from one gravitational potential to another, they must lose energy of an amount $h\nu\Delta\phi/c^2$. The loss in energy corresponds to a change in frequency such that

$$h\Delta\nu = h\nu\Delta\phi/c^2 \quad \text{or} \quad \Delta\nu/\nu = \Delta\phi/c^2.$$

Thus if an atomic clock were operated at sea level and signals sent from it to another identical clock at an altitude of 20,000 ft, the lower clock would appear to be slow by approximately 5 parts in 10^{-13}. This effect is small, but its detection and measurement would be of considerable interest.

[5] R. R. Howard and W. V. Smith, Phys. Rev. **79**, 128 (1950); C. H. Townes, Phys. Rev. **70**, 665 (1946).

[6] H. Margenau, Phys. Rev. **76**, 1423 (1949).

[7] J. M. Jauch, Phys. Rev. **72**, 715 (1947).

[8] A. S. Eddington, *Stars and Atoms* (Yale University Press, New Haven, Connecticut, 1927), p. 123.

SYNCHRONIZATION OF OSCILLATOR WITH SPECTRAL LINES

Having examined the characteristics of microwave absorption lines and their constancy under various conditions, we now consider the accuracy with which it is possible to synchronize a macroscopic electronic oscillation with that of the atomic or molecular system. This might be called the electrical engineering phase of the problem. However, only the fundamental limitations in accuracy of this synchronization will be discussed here, not the details and difficulties of particular circuits, even though for actual construction of an atomic clock these latter may be the more important.

Our problem can be specified by considering microwaves of a particular frequency transmitted through a wave guide filled with gas having an absorption line in the neighborhood of this frequency. The microwaves are detected after transmission. If a frequency change is to be detected, this change must correspond to a variation of absorption which results in a change in the detected signal larger than random fluctuations of thermal noise. Criteria must be developed for the wave guide dimensions, type of gas, gas pressure, microwave power, etc., which allow an optimum of this signal-to-noise ratio.

If the wave guide has a fractional loss per unit length α, and the gas absorption coefficient is γ, then of the microwave power P_0 introduced into a wave guide of length l, $P_0e^{-(\alpha+\gamma)l}$ will be transmitted. A change $\Delta\gamma$ in the gas absorption produces a change $\Delta P=-P_0le^{-(\alpha+\gamma)l}\Delta\gamma$ in the transmitted power. The change may be regarded as due to side-band signals of total power (when added in phase) $(\Delta P)^2/16P_0e^{-(\alpha+\gamma)l}$. This signal power must be comparable with, or greater than, the power $2kT\Delta f$ due to thermal radiation which is transmitted along the length of the wave guide and received in the detecting system. Here Δf is the band width of the signal amplifier and receiving system. Hence,

$$P_0l^2(\Delta\gamma)^2e^{-(\alpha+\gamma)l}/16\geq 2kT\Delta f.$$

The minimum detectable change in absorption coefficient is then

$$\Delta\gamma=(4/l)(2kT\Delta f/P_0)^{\frac{1}{2}}e^{(\alpha+\gamma)l/2}. \tag{9}$$

Differentiation of (9) with respect to l allows a determination of the optimum length of absorption cell, that is the length giving the smallest detectable change[9] $\Delta\gamma$

$$l_{0p}=2/(\alpha+\gamma). \tag{10}$$

Then

$$\Delta\gamma_{\min}=2e(\alpha+\gamma)(2kT\Delta f/P_0)^{\frac{1}{2}}. \tag{11}$$

The absorption coefficient of a line given by expression (2) may be written approximately as

$$\gamma=\frac{\gamma_{\max}(\Delta\nu)^2}{(\nu-\nu_0)^2+(\Delta\nu)^2}, \tag{12}$$

where $\gamma_{\max}$ is the absorption coefficient on the peak of the line, or when $\nu=\nu_0$.

Some of the devices used at present for setting on a spectral line set on the peak of the absorption. In that case, a change $\Delta\gamma$ in γ occurs if the frequency ν differs from ν_0 by an amount ϵ where

$$\epsilon=\Delta\nu\left(\frac{\Delta\gamma}{2\cdot\gamma_{\max}}\right)^{\frac{1}{2}}. \tag{13}$$

Considerably better performance can be obtained by using the steep slopes of the absorption lines, since then a small change in frequency corresponds to a greater change $\Delta\gamma$. For a distance from resonance $\nu-\nu_0=a\Delta\nu$,

$$(\epsilon/\nu)=(\Delta\gamma(a^2+1)^2/2a\gamma_{\max})(\Delta\nu/\nu).$$

Using (11), the smallest detectable fractional change in frequency is

$$\frac{\epsilon}{\nu}=\frac{e(\alpha+\gamma)(a^2+1)^2}{a\gamma_{\max}}\frac{\Delta\nu}{\nu}=(2kT\Delta f/P_0)^{\frac{1}{2}}. \tag{14}$$

For $\gamma_{\max}$ considerably smaller than α, the smallest value of ϵ/ν, or the steepest part of the resonant curve, is obtained when $a=\pm(\frac{1}{3})^{\frac{1}{2}}(\nu=\nu_0\pm(\frac{1}{3})^{\frac{1}{2}}\Delta\nu)$. Best performance for larger values of $\gamma_{\max}$ is obtained somewhat farther from the resonance frequency. For the other extreme $\gamma\gg\alpha$, optimum values of a are ±1. Thus for optimum a, the part of (14) which is written $(\alpha+\gamma)(a^2+1)^2/a\gamma_{\max}$ goes between the values $(16/3\sqrt{3})(\alpha/\gamma_{\max})$ when $\gamma_{\max}\ll\alpha$ and 2 when $\gamma_{\max}\gg\alpha$. For strong absorbers such as NH_3, it is possible to have $\gamma_{\max}>\alpha$, so that the minimum detectable frequency change given by (14) becomes

$$\epsilon/\nu=2e(\Delta\nu/\nu)(2kT\Delta f/P_0)^{\frac{1}{2}}. \tag{15}$$

The above expression would indicate that the error in frequency stabilization could be made arbitrarily small by making $\Delta\nu$ or T small or ν or P_0 large. It has already been mentioned however, that $\Delta\nu/\nu$ cannot be made smaller than the value $1/c(2kT\log_e2/m)^{\frac{1}{2}}$ given by the Doppler effect, and that T cannot be reduced indefinitely without reducing the gas pressure to too low a value. Usually an upper limit to the power P_0 is set by saturation effects, so that the only method of obtaining an arbitrarily small error is to decrease the band width Δf.

If a particular absorption line is considered, an optimum power flux can be determined. Equation (5) gives an apparent half-width when saturation occurs which, when inserted in place of the $\Delta\nu$ occurring in (15), gives

$$\frac{\epsilon}{\nu}=\frac{2e}{\nu}\left(1+\frac{8\pi^2|\mu|^2\nu It}{3ch\Delta\nu}\right)^{\frac{1}{2}}\left(\frac{2kT\Delta f}{P_0}\right)^{\frac{1}{2}}.$$

Since I, the number of quanta per second per unit cross-sectional area, is directly proportional to P_0, this ex-

[9] C. H. Townes and S. Geschwind, J. Appl. Phys. **19**, 795 (1948); W. Gordy, Revs. Modern Phys. **20**, 668 (1948).

pression approaches a limiting value for large P_0, and differs from the minimum value by only a factor of $\sqrt{2}$ when saturation first becomes noticeable, i.e., when

$$I \approx \frac{3ch(\Delta\nu)^2}{4\pi|\mu|^2\nu}.$$

If the cross-sectional area of the wave guide is A, then $P_0 = AhvI$, and for power P_0 large enough to give saturation,

$$\epsilon/\nu = (8e|\mu|/h)(kT\Delta f/3cA\nu^2)^{\frac{1}{2}}. \quad (16)$$

This expression very significantly does not contain the line width $\Delta\nu$. The advantage in sharpness of the line which is obtained by decreasing $\Delta\nu$ is just counteracted by the loss in sensitivity due to power saturation of the absorption. Thus there is no strong reason for attempting to attain lines so narrow that their width is determined by Doppler effects. In practice a very wide line is not desirable because the maximum power available P_0 and the power at which efficient detectors operate is usually near one milliwatt. In addition, nonfundamental variations of circuit response with frequency cannot be eliminated, and hence a sharp line is desirable to minimize their effects. However, Eq. (16) does show that so far as the limitations of thermal noise are concerned, the line width is unimportant, and can be adjusted to any convenient value. Since saturation of most spectral lines occurs with line widths of about one megacycle and a power flux of a few milliwatts per square centimeter, pressures such that the line is somewhat narrower than one megacycle would be a convenient choice. Of course, the radiation density or the field strength is not uniform throughout the wave-guide cross section nor throughout the length of the guide because of attenuation. However, a reasonably accurate approximation for ϵ/ν is obtained by considering the "average" saturation condition at the end of the guide where the radiation is introduced.

If the cross-sectional area of the wave guide is increased, the error given by (16) is reduced. However, it is generally not advantageous to increase the cross-sectional area of the wave guide to such a point that more than a few modes of the microwave radiation are propagated, because difficulties with the transmission properties of the wave guide and associated circuits are then encountered. Hence each cross-sectional dimension of the wave guide must be approximately equal to one wavelength λ, or $A \approx (c/\nu)^2$. We then obtain

$$\epsilon/\nu = (8e|\mu|/hc)(kT\Delta f/3c)^{\frac{1}{2}}. \quad (17)$$

For a numerical evaluation of (17), the specific case of the strongest ammonia line, the 3,3 line at 23,870 Mc, is of interest. Taking $T = 200°\text{K}$ and $\Delta f = 1$ cycle/sec, one obtains, since $\mu = 1.4\times10^{-18}$ esu,

$$\epsilon/\nu = 5\times10^{-15}. \quad (18)$$

This more or less fundamental limit in the frequency accuracy obtainable from absorption of microwaves in a gas during one second is thus considerably better than any present frequency stabilization.

It might be objected that an oscillator observed for only one second must have a spectral width of about 1 cycle/sec. For a frequency of 3×10^{10} cycles/sec, this would represent a fractional spectral width considerably greater than the accuracy given by (18). This accuracy refers only to the mean of the spectral distribution, and would still apply as long as the distribution is considerably smaller than the width of the absorption line. For longer times (smaller Δf) this apparent difficulty does not arise.

Even though actual accuracy obtained is likely to be limited for a long time by technical difficulties of a less fundamental nature than thermal noise, the expressions given above may be of some help in evaluating the importance of various parameters in determining accuracy. Expression (17) indicates that the ultimate accuracy is not dependent on the resonant frequency. It must be remembered, however, that this expression assumes losses in the wave guide are not larger than attenuation due to the gas and that an absorption path long enough to attenuate by a factor e^2 is assumed. In addition, power enough to saturate the lines is assumed to be usable. For low frequencies or weak lines a more general expression of the type (14) is more appropriate. However, since for ammonia the conditions validating (17) can easily be fulfilled, very little additional accuracy can be gained by use of any gas other than ammonia, except by decreasing $|\mu|$. Oxygen is one of the very unusual cases for which $|\mu|$ is smaller than for NH_3 by a factor of about 100 because the dipole moment is of magnetic rather than electric origin. For oxygen, it would be difficult to satisfy conditions validating (17). In addition, oxygen resonances would be somewhat disturbed by the earth's magnetic field due to the unusually large Zeeman effect mentioned above. However, if power sources and detectors near 5-mm wavelength are perfected, oxygen absorption lines may be useful for accurate frequency stabilization.

If a number N of "atomic clocks" stabilized on spectral lines are constructed and compared with each other and each has a fractional error ϵ/ν, the error of their average time, assuming the errors are of fundamental, and hence random origin, will be $\epsilon/\nu(N)^{\frac{1}{2}}$. Similarly from (14) or (17) if the band width is made narrower than the 1 cycle/sec chosen for the example calculated by a factor n, or the clock's time is averaged over a period of n seconds instead of 1 second, the expected fractional error is reduced by a factor $(n)^{\frac{1}{2}}$. Thus the theoretical fractional clock error decreases as the inverse square root of its running time, while the absolute value of the error increases continuously as the square root of this time.

USE OF DISPERSION

The above discussion has considered using only the absorption of a molecular resonance in order to stabilize an atomic clock. The dispersion, or reactive part of the resonance is equally usable. Figure 1 shows the behavior of the anomalous dispersion or variation in the dielectric constant near a resonant absorption. Near an absorption line give by (12), the dielectric constant of the absorbing material can be written

$$K=K_0+(\lambda(\nu_0-\nu)/\Delta\nu)\gamma$$
$$=K_0+\lambda(\nu_0-\nu)\gamma_{\max}\Delta\nu/[(\nu-\nu_0)^2+(\Delta\nu)^2] \quad (19)$$

where K_0 is the dielectric constant some distance on either side of the line, and other quantities are as defined for expression (12). For a gas at low pressure, K_0 may be taken as 1 and K is not very different from 1, so that

$$n=K^{\frac{1}{2}}=1+\lambda(\nu_0-\nu)\gamma/2\Delta\nu. \quad (20)$$

In order to use the dispersion to give a signal, some microwave circuit must be used which detects the change in phase of the signal. If the microwave voltage appearing at the output of a wave guide of length l is $V_0\exp(-\alpha l/2)\sin 2\pi\nu t$, where α is the attenuation constant of the guide as before and V_0 is the input voltage, then from (20) the output voltage when the guide is filled with gas is

$$V=V_0e^{-(\alpha+\gamma l/2)}\sin\left[2\pi\nu t-\frac{\pi(\nu_0-\nu)}{\Delta\nu}\gamma l\right]$$
$$=V_0e^{-(\alpha+\gamma l/2)}\left[\sin 2\pi\nu t\cos\pi\frac{(\nu_0-\nu)\gamma l}{\Delta\nu}\right.$$
$$\left.+\cos 2\pi\nu t\sin\pi\frac{(\nu_0-\nu)}{\Delta\nu}\gamma l\right]. \quad (21)$$

The out-of-phase signal due to the gas dispersion is

$$S=V_0e^{-(\alpha+\gamma l/2)}\sin\frac{\pi(\nu_0-\nu)}{\Delta\nu}\gamma l\cos 2\pi\nu t.$$

This signal can be shown to change most rapidly with frequency at the steepest slope of the dispersion curve, i.e., at $\nu_0=\nu$. In that case

$$\Delta S=\frac{V_0\pi\gamma_{\max}l}{\Delta\nu}e^{-(\gamma_{\max}+\alpha)l/2}\epsilon,$$

where ϵ is the change in frequency $\nu_0-\nu$. The error signal ΔS is a maximum with respect to guide length l when $(\alpha+\gamma_{\max})l=2$, which is very much like the optimum length for absorption, given by $(\alpha+\gamma)l=2$. Hence

$$\Delta S=\frac{V_0 2\pi\gamma_{\max}e^{-1}}{\gamma_{\max}+\alpha}(\epsilon/\Delta\nu). \quad (22)$$

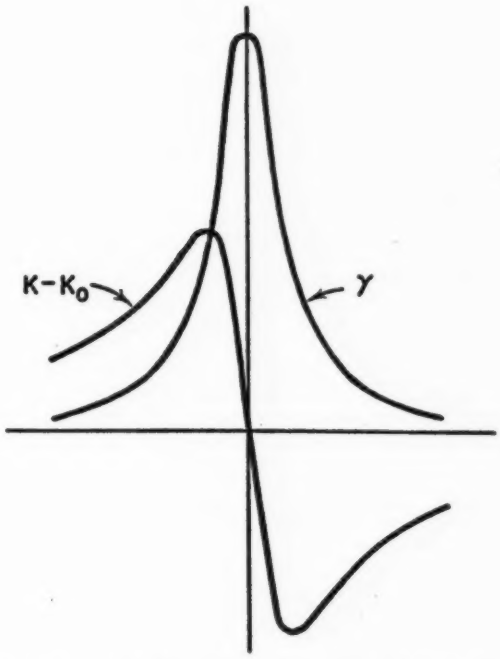


FIG. 1. Loss and dispersion near a spectral line. If $\gamma_{\max}$ is written for the peak value of the absorption coefficient γ, then the variation in the dielectric constant $K-K_0$ between maximum and minimum is $\lambda\gamma_{\max}$, where λ is the wavelength.

The power in this signal is proportional to $(\Delta S)^2$, and may be written

$$\Delta P=P_0\left(\frac{2\pi\gamma_{\max}e^{-1}}{\gamma_{\max}+\alpha}\frac{\epsilon}{\Delta\nu}\right)^2$$

where P_0 is the initial power introduced into the wave guide. If this power is equal to thermal noise $kT\Delta f$, then the fractional frequency error is

$$\epsilon/\nu=((\alpha+\gamma_{\max})e\Delta\nu/2\pi\gamma_{\max}\nu)(kT\Delta f/P_0)^{\frac{1}{2}}. \quad (23)$$

Use of dispersion, therefore, gives not only essentially the same optimum length for the microwave path in gas, but also within a small factor the same fractional frequency error as is shown by comparing (23) and (14).

USE OF ATOMIC OR MOLECULAR BEAMS

Spectroscopic lines detected in molecular or atomic beams are also of great interest as frequency standards. For this type of spectroscopy, a beam of atoms or molecules usually travels through a region in which there is a radiofrequency field. If the field is of proper frequency, some particles in the beam undergo a transition, after which they may be separated from particles which have not made a transition, and then detected. Very sharp resonances can be obtained because the field can interact with the particles over a comparatively long time, and may be oriented with respect to the particle velocities so that Doppler effects are essentially eliminated.

If a radiofrequency field interacts with particles in a beam during a time t (not necessarily continuously), then if a resonant response occurs, it may be as sharp[10] as $\Delta\nu=1/t$. For atoms of thermal velocities (approximately 10^5 cm/sec) and a beam of length 100 cm, a resonance as sharp as 1 kilocycle is hence obtainable.

During the travel of the beam, transitions may be induced not only by the applied radio or microwave field which is to be stabilized, but also by random fields associated with thermal radiation. If the region of interaction is considered a wave guide, then power kT is flowing down the wave guide in each mode of propagation due to thermal fluctuations. This gives a fluctuating electric field strength of approximately

$$E=(1/\lambda)(kT\Delta f/c)^{\frac{1}{2}}. \tag{24}$$

The fraction of the beam of particles in which transitions will be induced due to this thermal radiation is $F=E|\mu|t/h$, where $|\mu|$ is the dipole matrix element for the transition and t is the time of interaction. These thermally induced transitions produce a fractional uncertainty in the frequency of stabilization which is roughly $\epsilon/\nu=\Delta\nu F/\nu$, which becomes, after substituting for F and E,

$$\epsilon/\nu=(\Delta\nu/\nu)(|\mu|t/h\lambda)(kT\Delta f/c)^{\frac{1}{2}}.$$

Remembering that $t\approx 1/\Delta\nu$ and $\nu\lambda=c$,

$$\epsilon/\nu=(|\mu|/hc)(kT\Delta f/c)^{\frac{1}{2}}. \tag{25}$$

This expression is essentially the same as (17), which gives the limiting accuracy using the microwave absorption technique.

Although expressions (25) and (17) are similar, it must be pointed out that molecular or atomic beam techniques offer rather definite means of avoiding trouble from the thermodynamic limit as derived above.

[10] N. F. Ramsey, Phys. Rev. 78, 695 (1950).

It is a simple matter to send the beam through a region where the various modes of thermal disturbances produce only weak fields. Hence the approximate formula (24) for the fluctuation field would give much too large a value for the field E. The applied field in such an arrangement would be relatively weak also, but may almost always be increased to a value large enough to give the desired effect on the beam.

Another source of fluctuation in frequency stabilization with molecular beams is a fluctuating beam intensity. Experimentally, rather large fluctuations of the beam are sometimes observed due to poor vacua, acoustic vibrations, detector noise, and other nonfundamental but still important difficulties. More fundamental fluctuations occur because of the finite number of particles in the beam. If the beam current is I, then the number of particles in time t_0 is It_0/e, and the fractional error in frequency produced by fluctuation in the number of particles is

$$\epsilon/\nu=(\Delta\nu/\nu)(e/It_0)^{\frac{1}{2}}$$

or

$$\epsilon/\nu=(\Delta\nu/\nu)(e\Delta f/I)^{\frac{1}{2}}, \tag{26}$$

where $\Delta f\approx 1/t_0$ is the effective band width. Usually the limit set by (26) would be considerably less than that given by (25).

Although molecular or atomic beam techniques are not easy, especially for reliable long-term operation, they offer the possibility of a high accuracy of stabilization. Their most valuable characteristic is the very sharp resonance or high Q which can be obtained and which can considerably simplify many of the practical difficulties in approaching the theoretical limit of accuracy described here.

ACKNOWLEDGMENT

The above discussion has been stimulated by the interest and encouragement of Dr. Harold Lyons.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 11 NOVEMBER, 1951

The Nature of the Static and Kinetic Coefficients of Friction

ERNEST RABINOWICZ
Lubrication Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received May 23, 1951)

Experiments have been carried out to determine the transition between static and kinetic conditions when stationary metal surfaces are set into motion, a simple method being used which measures the energy that has to be given to one of the bodies to start it moving. The method is confined to cases in which the static coefficient exceeds the kinetic. Using a load of 1 kg and metal surfaces of various kinds, it is found that the static coefficient persists for distances of the order of 10^{-4} cm, and then gradually falls off to values corresponding to the kinetic coefficient. This behavior is shown to be consistent with a simple model based on the assumption that the friction force is needed to shear metallic junctions formed between the metal surfaces. The action of boundary lubricants is discussed, and it is shown that they can act either by diminishing the metallic interaction directly, or by preventing its increase during the sliding process.

IT has long been known that the force to set into motion two surfaces at rest is often greater than the force to maintain motion. For such surfaces it is usual to quote two values of the coefficient of friction; μ_s, the static coefficient, which corresponds to the force to start motion, and μ_k, the kinetic coefficient, operative for moving surfaces.

The frictional behavior of these surfaces is often represented on a μ-v diagram, in which the coefficient of friction is plotted as a function of the velocity of sliding (Blok).[1] Figure 1 shows the simplest case, namely a constant kinetic coefficient of friction at all sliding velocities and a higher static coefficient, and this appears to fit the case of the sliding of two clean surfaces consisting of different metals (Sampson, Morgan, Reed, and Muskat).[2] Figure 2 shows a case common for poorly lubricated surfaces (Bristow).[3] It must be emphasized that there has been little investigation of such curves; it is probable that, at sufficiently low velocities of sliding, the curve shown in Fig. 1 will be found to contain the gradually falling characteristic of Fig. 2.

However, curves such as those shown in Figs. 1 and 2 do not describe fully the behavior of the surfaces. For, suppose the surfaces with frictional properties described in Fig. 1, initially at rest, are suddenly set into motion with a constant velocity V. The initial force resisting motion is that corresponding to μ_s; after sliding has proceeded for some distance, it is μ_k. Figure 1 tells nothing of the transition between these two conditions.

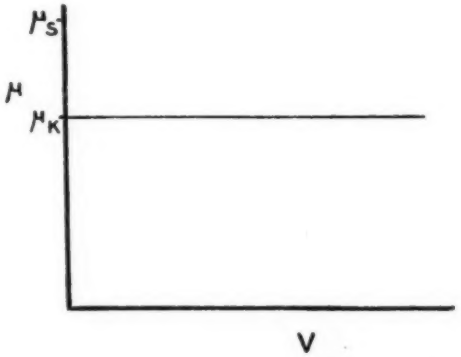


FIG. 1. Possible variation of coefficient of friction with sliding velocity.

Such information is given by a μ-s curve, which plots the coefficient of friction against distance traveled. As a first approximation we may assume that the static coefficient operates for a distance s_k, when a rapid drop to kinetic values takes place (Fig. 3). It will be assumed further that the curve is the same over a wide range of speeds of sliding.

The present paper describes a method for determining s_k, the distance over which the static coefficient operates, and a slight modification of the experimental procedure allows detailed investigation of the shape of the μ-s curve.

THEORY UNDERLYING THE METHOD

A block of mass M rests on a plane inclined at an angle θ (Fig. 4), the angle θ being chosen so that $\tan\theta$ is just greater than μ_k for the block on the plane, but considerably less than μ_s, i.e.,

$$\mu_s \gg \tan\theta > \mu_k; \tag{1}$$

thus, if the block M is in motion down the plane, it

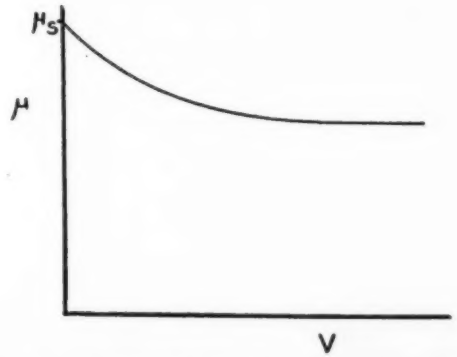


FIG. 2. Another possible variation of friction coefficient with sliding velocity.

[1] H. Blok, J. Soc. Automotive Engrs. **46**, 54 (1940).
[2] Sampson, Morgan, Reed, and Muskat, J. Appl. Phys. **14**, 689 (1943).
[3] J. R. Bristow, Proc. Roy. Soc. (London) **A189**, 88 (1947).

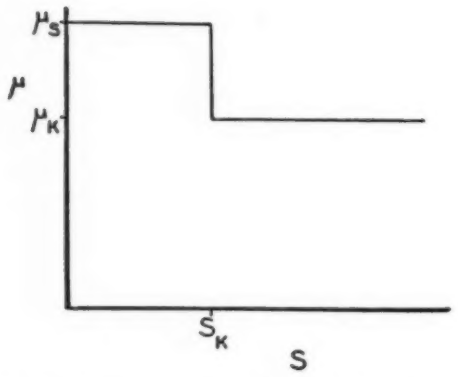


FIG. 3. Idealized representation of the variation of coefficient of friction with distance.

will remain in motion, since

$$Mg\sin\theta > \mu_k Mg\cos\theta. \tag{2}$$

If M is at rest, it will remain at rest, since

$$Mg\sin\theta < \mu_s Mg\cos\theta. \tag{3}$$

A small ball of mass m, initially at rest, is rolled a distance x down the plane and impacts the stationary block M. It is found that there is a critical distance l_k such that for $x > l_k$, the block is set into motion right down the plane, whilst for $x < l_k$, the block appears to remain at rest—though, as will be seen, microscopic motion may occur.

It is then possible to relate the distance l_k with the distance s_k, the limit of the persistence of the static coefficient of friction.

We may divide the impact and its consequences into two parts: (a) Transfer of momentum from m to M, (b) Motion of M.

(a) Transfer of Momentum

The velocity v of a ball sliding a distance l_k down the plane in the absence of friction is given by

$$v^2 = 2\cdot g l_k \sin\theta. \tag{4}$$

For a rolling ball, Eq. (4) must be modified since the ball acquires rotational as well as translational energy. A simple calculation shows that in this case

$$v^2 = 10/7\cdot g l_k \sin\theta. \tag{5}$$

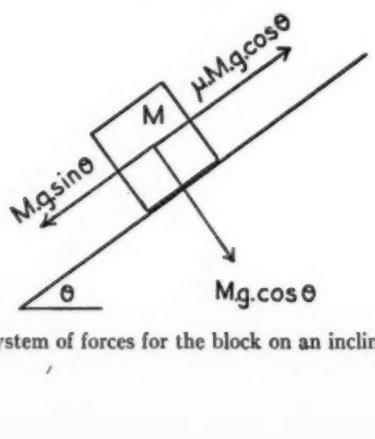


FIG. 4. System of forces for the block on an inclined plane.

The velocity V of the block after impact is given by

$$MV = (1+e)mv, \tag{6}$$

where e is the coefficient of restitution, and $M \gg m$.

(b) Movement of the Block

The force opposing motion down the plane is initially

$$\mu_s Mg\cos\theta - Mg\sin\theta.$$

This will bring the block to rest in a distance X given by

$$\begin{aligned} V^2 &= 2g(\mu_s\cos\theta - \sin\theta)X \\ &= 2g\cos\theta(\mu_s - \tan\theta)X. \end{aligned} \tag{7}$$

If $X < s_k$, the block will actually be brought to rest. If $X > s_k$, then before the block comes to rest the decelerating force $Mg\cos\theta(\mu_s - \tan\theta)$ becomes the accelerating force $Mg\cos\theta(\tan\theta - \mu_k)$, and the block moves right down the plane.

The limiting condition is given by $X = s_k$. Combining

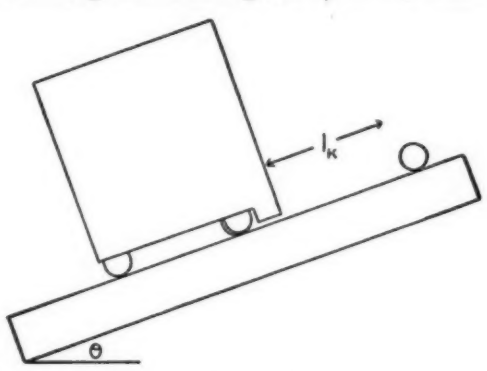


FIG. 5. Sketch of the apparatus.

Eqs. (5)–(7) we have

$$s_k = \frac{5}{7}\cdot(1+e)^2\cdot\frac{m^2}{M^2}\cdot\frac{\tan\theta}{\mu_s - \tan\theta}\cdot l_k. \tag{8}$$

EXPERIMENTAL

Experiments were at first carried out using clean copper sliding on steel. The block M consisted of a steel block 2-in. cube with an additional ledge along one edge of the bottom face (Fig. 5). Into this face were screwed 3 copper hemispherically ended sliders, equilaterally spaced. When the block was placed on the inclined plane, most of its weight was borne by one slider. The bottom surface was a stout steel strip, which could be tilted about its middle.

Prior to the experiments, the steel strip was cleaned by lapping on emery paper under water, then dried on clean filter paper. The copper sliders were cleaned by being rotated in the shaft of a motor while clean emergy paper was pressed against them.

For the banging experiments, a hardened steel ball of $\frac{3}{8}$-in. diameter was used. The coefficient of restitu-

tion e, determined by measuring the height of rebound after collision with the ledge of the block, was found to be 0.5.

For copper on steel, $M=1000$ g, μ_s was measured from the angle of tilt to set into motion a stationary block in the absence of any lateral force, and was found to be 0.46. μ_k, the tangent of the smallest angle of tilt which would maintain the block in motion, was measured as 0.31.

When the strip was set at an angle such that $\tan\theta$ was 0.33, it was found that l_k, the critical distance, was 2.0 cm.

s_k can then be calculated from Eq. (8) as 1.0×10^{-4} cm.

Note.—This calculation is valid only if the time of collision is shorter than the time to bring the block to rest, as otherwise some of the momentum imparted by the ball will be wasted. The time to bring the block to rest is readily calculated to be 1.5×10^{-3} sec. The time of collision may be estimated at 2.10^{-4} sec (Bowden and Tabor).[4] Thus the above treatment is justified.

Experiments were also carried out with other metal combinations, and the results obtained are summarized in Tables I and II.

TABLE I. Clean metal surfaces. Mass of block = 1 kg.

Metals used	μ_s	μ_k	$\tan\theta$	s_k
Copper on mild steel	0.46	0.31	0.33	1.0×10^{-4} cm
Lead on mild steel	0.72	0.47	0.50	3
Mild steel on copper	0.54	0.39	0.41	0.9
Mild steel on titanium	0.63	0.45	0.47	6
Mild steel on zinc	0.65	0.47	0.49	2

From Table I it will be seen that copper on steel and steel on copper yield similar values of μ and s_k. With steel as the other metal, use of the softer lead leads to a greater value of s_k than does that of the harder copper. Zinc, with intermediate hardness, has an intermediate value of s_k.

Copper on copper, clean, gave a value of μ_s too high to be measured by the apparatus used. It is known (Bowden and Leben)[5] that μ_k is also very high, and since μ_s and μ_k are equally large this combination cannot be studied by this method. Similarly with stearic acid, the static and kinetic coefficients are equal, and again s_k is indeterminate. In the case of the surfaces lubricated by cetane there is a complication caused by the fact that the μ-v curve is likely to be that shown in Fig. 2 rather than the simple one of Fig. 1. A numerical correction is therefore applicable to the corresponding values of s_k shown in Table II, but the correction is not too large and it has not seemed worth while investigating this point fully.

[4] F. P. Bowden and D. Tabor, Engineer (London) **172**, 380 (1941).

[5] F. P. Bowden and L. Leben, Proc. Roy. Soc. (London) **A169**, 371 (1939).

TABLE II. Surfaces in various stages of lubrication.

Surfaces used	μ_s	μ_k	$\tan\theta$	s_k
Copper on copper				
Clean	>1.0	>1.0	—	—
With pure cetane	0.58	0.40	0.42	1.2
With slightly impure cetane	0.35	0.23	0.25	1.3
With solid stearic acid	0.11	0.11	—	—
Mild steel on copper				
Clean	0.54	0.39	0.41	0.9
With pure cetane	0.32	0.21	0.225	0.7
With solid stearic acid	0.10	0.10	—	—
Mild steel on titanium				
Clean	0.63	0.45	0.47	6
With pure cetane	0.61	0.45	0.47	5
Soln. of stearic acid in cetane	0.44	0.29	0.31	4
With solid stearic acid	0.20	0.15	0.16	2

The results with titanium are anomalous. Titanium is harder than copper, but has a larger value of s_k than even lead. Its friction with stearic acid is also rather large.

DETERMINATION OF THE TRUE SHAPE OF THE μ-s CURVE

If observations of the critical ball distance l are made for various angles of tilt θ, it is possible to determine the shape of the μ-s curve.

Equation (8) is equivalent to

$$s_k(\mu_s-\tan\theta)=\frac{5}{7}\cdot(1+e)^2\cdot\frac{m^2}{M^2}\cdot l_k\tan\theta \qquad (9)$$

where

$$K=5/7\cdot(1+e)^2\cdot m^2/M^2.$$

Now the quantity $s_k(\mu_s-\tan\theta)$ has the dimensions of area on a μ-s diagram; in Fig. 3 it represents the area between the μ-s curve and the line $\tan\theta=$ constant (for $\mu_k<\tan\theta<\mu_s$). If the actual μ-s curve is more general than that shown, we must write this area in for form of an integral, namely

$$\int_{\tan\theta}^{\mu_s} s\cdot d\mu. \qquad \text{(See Fig. 6.)}$$

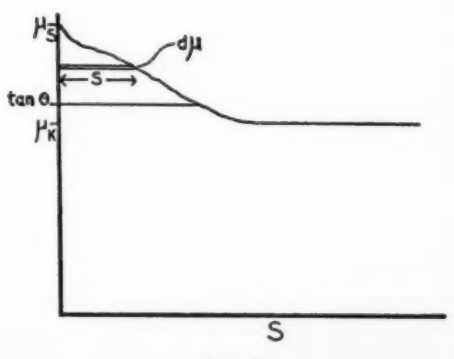


FIG. 6.

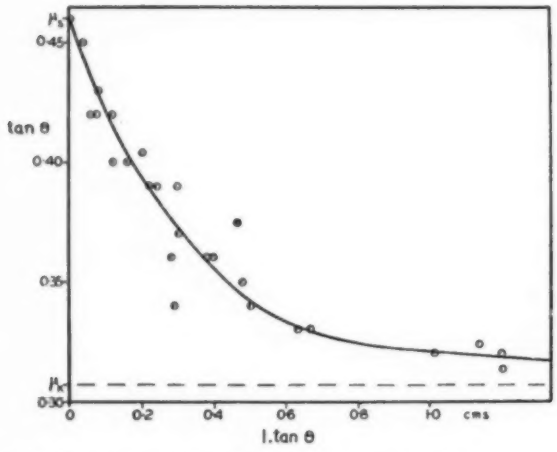


FIG. 7. Tanθ plotted as a function of l.tanθ for clean copper on mild steel, mass of $M=1000$ g, mass of $m=3.5$ g.

Thus we have

$$\int_{\tan\theta}^{\mu_s} s\cdot d\mu = K\cdot l\cdot\tan\theta \qquad (10)$$

or

$$s = K\cdot d/d\mu(l\cdot\tan\theta). \qquad (11)$$

Experimentally, the tilt of the plane is varied, and l, the critical ball distance, determined for each inclination. Tanθ is plotted as a function of $l\cdot\tan\theta$ and the differential of this curve gives the μ-s curve (tanθ being equivalent to μ for our experimental arrangement).

Figure 7 shows a set of experimental points for copper on steel, each point plotted being the average of five determinations. It is seen that the points show considerable scatter, which is due to local variations of a few percent in the value of μ such as almost always exist with metal surfaces, rather than real variations by factors of two or more in the values of l.

Figure 8, calculated as the differential of the curve in Fig. 7, is the required μ-s curve.

μ-s curves have also been obtained for the other metal combinations studied and it is found that the various curves have nearly the same shape.

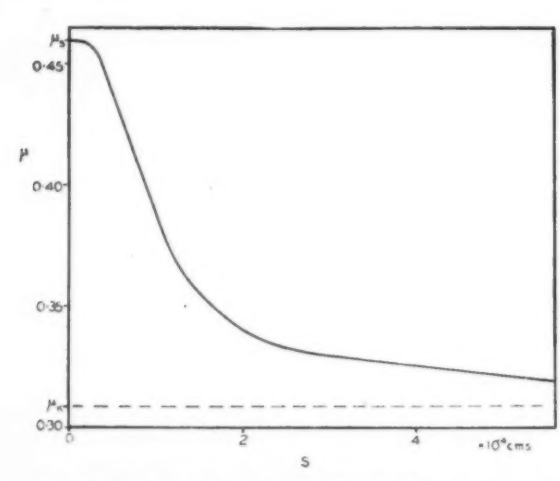


FIG. 8. Differential of the curve of Fig. 7, showing variation of the coefficient of friction with distance.

THE ELASTIC PART OF THE μ-s CURVE

Previous investigations by a number of workers have shown that, before sliding commences, an elastic displacement of the contacting surfaces takes place. Stevens,[6] Rankin,[7] and Haikin, Lissovsky, and Solomonovitch,[8] working with various surface conditions, have determined this elastic limit to be of the order of $5\cdot10^{-6}$ cm. Both Rankin and Stevens found the elastic region to extend to stresses corresponding to the coefficient of static friction in their experiments. This result is important to the present investigation in that it removes uncertainty as to the shape of the μ-s curve shown in Fig. 8.

Owing to the fact that our determinations of this curve have consisted essentially of measurements of the area bounded by the curve in Fig. 7 and various horizontal lines, it is impossible to distinguish between the possible μ-s curves shown in Figs. 9 and 10, or others of this class, all of which have the same value of s for any given value of μ. For since all such curves give equal magnitudes to the quantity

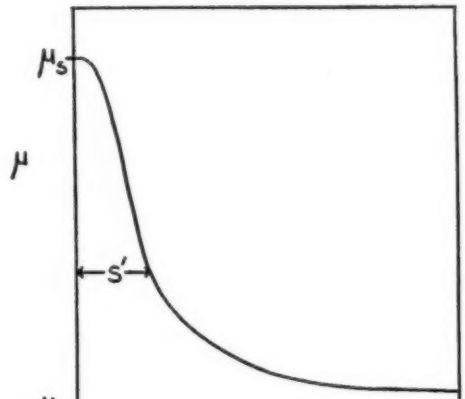


FIG. 9. Possible form of the μ-s curve consistent with the data of Fig. 7.

$$\int_{\tan\theta}^{\mu_s} s\cdot d\mu$$

which appears in Eq. (10), they will all give rise to the same curve of tanθ as a function of $l\cdot\tan\theta$ in Fig. 7. Hence both Figs. 9 and 10 are consistent with the data of Fig. 7.

To obtain an unequivocal μ-s curve we must make a further assumption, that the first part of the deformation, the elastic stage, extends right up to the full

[6] J. S. Stevens, Phys. Rev. **8**, 49 (1899).
[7] J. S. Rankin, Phil. Mag. **2**, 806 (1926).
[8] Haiken, Lissovsky, and Solomonovitch, J. Phys. **1**, 455 (1939).

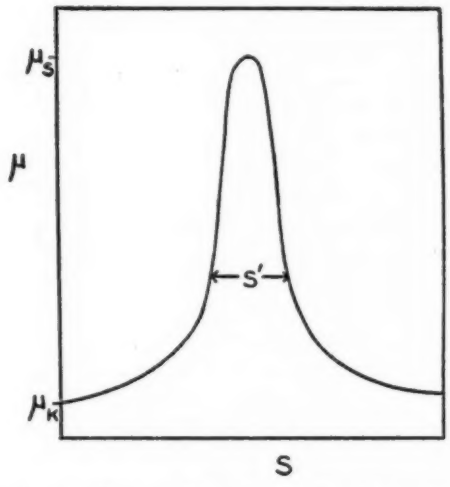


FIG. 10. Another possible μ-s curve equally consistent with the data of Fig. 7.

value of the static coefficient of friction. Figure 11 may then be drawn, and the full line shows the final curve.

RELATION BETWEEN THE μ-s CURVE AND THE SIZE OF THE METALLIC JUNCTIONS

Bowden, Moore, and Tabor[9] have shown that, when clean metal surfaces are in contact, metallic junctions are formed between them, and they are sheared during sliding. Since the μ-s curve is a force-extension curve, it is of interest to see if it can be interpreted as the force-extension curve of these metallic junctions. To do this, some assumptions must first be made as to the nature of the metallic junctions and their mode of deformation.

To get a simple picture, it will be assumed that all the metallic junctions are circular areas of diameter d, and that sliding produces a simple shearing along the interface between the surfaces. The force resisting shear at any instant will be assumed proportional to the real area of contact at that instant (Fig. 12).

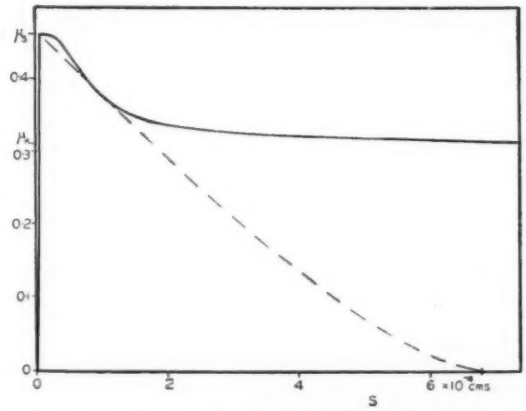


FIG. 11. Final μ-s curve (full line). The curve shown in Fig. 13 drawn so as to be tangential to the μ-s curve (broken line).

[9] Bowden, Moore, and Tabor, J. Appl. Phys. **14**, 80 (1943).

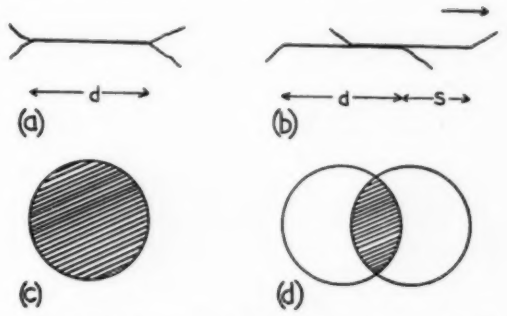


FIG. 12. Idealized representation of a typical metallic junction in elevation and plan, the shaded region being the true area of contact. (a) and (c):—as formed during static loading. (b) and (d):—after sliding a distance s.

On this assumption, the frictional force may be calculated by simple geometry as a function of s, the distance moved, and a plot is shown in Fig. 13. This curve is also shown in Fig. 11, and it will be seen that in some ways it resembles the μ-s curve observed. The main difference lies in the fact that, in actual practice, as soon as sliding commences the breaking of the original metallic junctions is to some extent compensated for by the formation of new ones. These new junctions are not as strong as the original ones and hence the kinetic coefficient is less than the static, but, since it is not zero, the μ-s curve does not go down to the abscissa.

A second difference is that the μ-s curve commences with a nearly horizontal portion. It appears that the decrease in area of each junction as sliding commences is compensated for by the formation of stronger welds over the area that remains, perhaps by a work-hardening process. Another contribution to this effect arises from the fact that not all the metallic junctions shear in the simple way shown in Fig. 12. In a small proportion of the cases the shearing takes place within the bulk of one of the metal asperities (Fig. 14(a)) and for these junctions there is almost certainly a higher friction force at small values of s than with the others.

The most satisfactory basis for comparison of the curves shown in Figs. 11 and 13 consists in making the

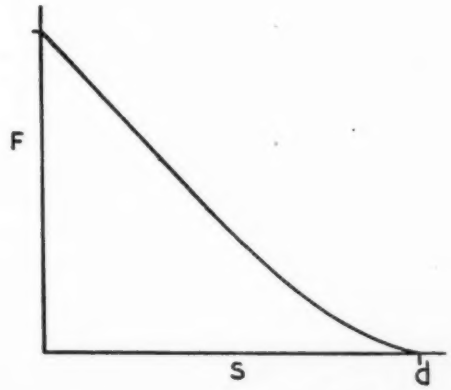


FIG. 13. Shearing force as a function for s for the junction shown in Fig. 12.

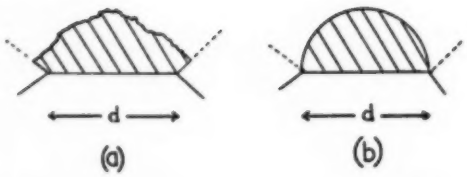


FIG. 14. (a) Representation of a typical metallic fragment transferred from the top to the bottom surface during sliding. (b) Idealized model of (a).

F-d curve tangential to the μ-s curve (this is represented by the broken line in Fig. 11). A value of d is obtained of $7 \cdot 10^{-4}$ cm (copper on steel, normal component of the load $[M \cos\theta] = 900$ g).

This value is to be compared with one obtained from measurements on the size of metal fragments transferred between sliding surfaces. As was mentioned above, the break between the metal junctions sometimes takes place within one of the metals, not at the interface. As a result, a metal fragment is transferred from one surface to the other.[9] Rabinowicz and Tabor,[10] using an autoradiographic technique, have measured the size of the fragments transferred in the sliding of copper on steel, load 2 kg, and have found an average mass of 10^{-8} g. Figure 14(a) represents a possible shape for the typical fragment, and Fig. 14(b) shows a simplified model, in which the fragment is considered as a hemisphere, its plane surface being the circle of contact of the junction. Using this model, d is calculated as $17 \cdot 10^{-4}$ cm. In view of the many complicating factors involved, the agreement of this figure with the $7 \cdot 10^{-4}$ cm previously obtained may be considered very satisfactory.

In theory, an estimate of the average junction size can be obtained by examining the distance at which the actual value of μ first becomes equal to μ_k. This distance should be closely related to the average junction size. In practice the actual μ-s curve becomes nearly asymptotic to the value of μ_k, and accurate measurements cannot be made, but the results obtained are of the right order of magnitude and tend to provide confirmation of the model presented.

DISCUSSION

Experiments have been carried out which lead to a determination of the curve of friction plotted as a function of distance travelled for surfaces which have been in stationary contact for times of the order of seconds, and are then set into motion. This curve is a characteristic feature of the frictional properties of the surfaces, but has been little investigated previously.

It has been suggested by Bowden, Moore, and Tabor[9] that, when metal surfaces are in contact, plastic flow takes place and metallic junctions are formed between them. Qualitative support for this view is provided by the fact that, in our experiments, the static coefficient of friction persists for greater distances in the case of lead on steel than with copper on steel. For, it is highly probable that this persistence of μ_s is directly related to the size of the metallic junctions, and the softer lead may be expected to give rise to larger junctions than the harder copper.

In fact it is possible to develop a quantitative relationship. Using a simple model in which the metallic junctions are considered as circular regions of equal size, and sliding is assumed to occur as the shearing of these junctions along their interface, it is possible to calculate the size of the metallic junctions from the shape of the μ-s curve, and the value obtained is in good agreement with values derived by another method involving the size of metal fragments.

The reason for the greater static coefficient of friction compared with the kinetic must now be considered. Sampson, Morgan, Reed, and Muskat[2] found that surfaces stationary for very short times gave low (kinetic) values of the coefficient of friction. These and other experiments suggest that the strength of the metallic junctions depends on the time of stationary contact of the surfaces, being smaller for times of the order of milliseconds than for times of the order of seconds. Sliding surfaces are equivalent to surfaces in stationary contact for very short times, and will give rise to low coefficients of friction.

It seems that this greater strength of junctions formed after long contact is due to a process analogous to creep. As a result, the junctions increase in size with time, and the force to shear these larger junctions is greater than the force which would have been required initially.

The process of sliding is envisaged as follows: when the tangential load is applied, at first elastic deformation of the asperities and the substrate takes place, and continues until the shear strength of the junctions is reached. Shearing of the junctions now takes place and the coefficient of friction falls off as the strong junctions which were formed during static loading become replaced by weaker ones. The influence of the strong junctions persists over a distance that is simply related to the average junction size.

It is of interest to consider briefly the cases to which the method is inapplicable, those of clean surfaces of the same metal and well-lubricated surfaces of the same or different metals. With clean, like metals the junctions are very strong, and shearing of the junctions is accompanied by severe distortion of the surfaces. It seems plausible that this distortion brings into contact parts of the surface which otherwise would have remained separate and thus, by a "snowballing" mechanism, the true area of contact and therefore the resistance of the surfaces to motion become greater as sliding proceeds. This effect was first noted in the case of copper by Ernst and Merchant.[11] The anomalous

[10] E. Rabinowicz and D. Tabor, Proc. Roy. Soc. (London), to be published.

[11] H. Ernst and M. E. Merchant, Conf. Friction and Surface Finish M.I.T. (1940), p. 76.

frictional properties of titanium are probably also due to a "snowballing" effect. This would explain the abnormally large metallic junctions.

With lubricated surfaces further complications are met with due to the properties of the lubricants. It is of advantage to consider separately the cases of well-lubricated and poorly lubricated surfaces. In the case of well-lubricated surfaces, the metallic junctions play only a small part in determining the friction coefficient, since the metallic junctions are present over only a very small part of the real area of contact (Rabinowicz and Tabor).[10] In the case of copper surfaces lubricated by palmitic acid, a low value of the coefficient of friction (~ 0.1) has been found with μ_s no greater than μ_k. It is suggested that these values are properties of the lubricant rather than those of the metallic junctions.

With poorly lubricated surfaces, however, the metallic junctions make an important contribution to the total frictional resistance. The frictional properties of copper on copper, lubricated with cetane, are similar to those of clean copper on steel, both the coefficient of friction and the size of the junctions being approximately the same. The main function of the cetane has been to prevent the "snowballing" action by which, in the case of clean copper on copper, the area of contact and consequently the friction increase as sliding proceeds. The primary metallic interaction has also been reduced, but still contributes substantially to the total frictional force.

Confirmation of this explanation is seen in the case of unlike metals lubricated with cetane. The cetane produces less reduction in the coefficient of friction than with like metals, since even in its absence the tendency for "snowballing" to occur is small.

Thus we may consider a lubricant as having two possible functions; either, to prevent the formation of metallic junctions between the surfaces, more accurately, greatly to decrease the fraction of the real area of contact that consists of metallic junctions, and this is achieved by good lubricants; or, to prevent the "snowballing" process by means of which metallic junctions grow during sliding. Even poor lubricants can often achieve this. In this latter case the metallic interaction may be quite large, but surface damage and friction are comparatively small.

SUMMARY AND ACKNOWLEDGMENTS

The variation of the value of the coefficient of friction with distance of sliding has been measured by a simple method in which a small sphere is impacted against a large block on an inclined plane. The method is limited to those cases in which the static coefficient is greater than the kinetic.

The results show that the static coefficient persists for distances of the order of 10^{-4} cm, the exact distance depending on the nature of the surfaces, and then falls off gradually until the kinetic coefficient is reached at about 10^{-3} cm.

A simple model is developed in which the force resisting motion is taken to be the force to shear the metallic junctions formed between the metal surfaces in contact, and a value for the size of the metallic junctions is calculated which is in good agreement with results obtained by an autoradiographic method.

The static coefficient will normally exceed the kinetic, since the metallic junctions become stronger after the surfaces have been in stationary contact for some time. Exceptions to this rule are explained, in the case of clean like metals, by a "snowballing" effect as a result of which the true area of contact increases during sliding; in the case of well-lubricated surfaces by the fact that the friction is determined almost exclusively by the shear properties of the lubricant layer.

It is with pleasure that I thank Professor J. T. Burwell, Jr., for his encouragement and interest, Dr. D. Tabor and Mr. E. Eisner of Cambridge University for helpful discussions, and the Chrysler Corporation for financial support.

The Crystallographic Aspect of Slip in Body-Centered Cubic Single Crystals. I. Theoretical Considerations

A. J. OPINSKY* AND R. SMOLUCHOWSKI
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received July 9, 1951)

The mechanism of slip in body-centered cubic lattices is considered in terms of the relation between the orientation of the tensile axis and the active slip system. A "projection" of points in the unit stereographic triangle on its [001]–[011] side permits an analysis of experimental data and leads to an accurate determination of the ratio of the critical shear stresses on various planes.

PROPERTIES of binary solid solutions are often affected by changes in the distribution of the two kinds of atoms in the crystalline lattice. In particular it has been suggested that a deviation from randomness which leads to the formation of oriented clusters[1] of one kind of atoms may influence certain anisotropic phenomena, such as slip in body-centered cubic lattices, precipitation, etc. In body-centered cubic lattices the active slip system depends strongly upon temperature and upon composition, and this effect may be related to the influence of clustering on the critical shear stress of various slip planes. Besides the temperature dependence of slip systems in pure body-centered cubic metals,[2] the best known example of the influence of temperature and of composition on slip in an alloy is provided by the iron-silicon alloys studied by Barrett, Ansel, and Mehl,[3] as illustrated in Fig. 1. The present paper summarizes the general analysis of the slip behavior of body-centered cubic lattices. In a later paper results of experiments on iron-silicon alloys, and a reinterpretation of data available from the literature will be given.

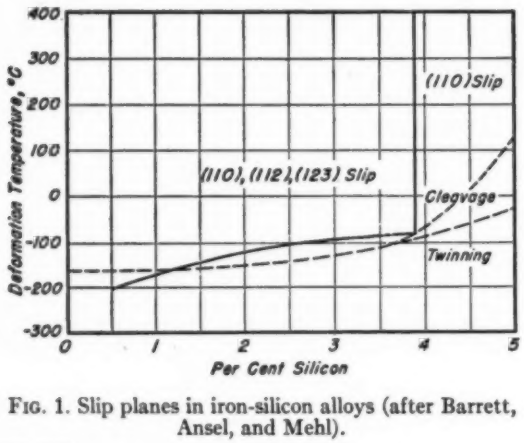


FIG. 1. Slip planes in iron-silicon alloys (after Barrett, Ansel, and Mehl).

1. THE ORIENTATION DEPENDENCE OF THE RESOLVED SHEAR STRESS

Despite numerous investigations, the actual mechanism of slip as a function of the orientation of the body-centered cubic lattice has not been analyzed successfully. Fahrenhorst and Schmid,[4] in a comprehensive treatment of the slip behavior of single crystals of iron, did not consider the possibility of operation of several different types of planes, but tried to determine the one type of slip plane which they thought operated. Gough,[5] using a different state of stress, that of alternating torsion, concluded that the operating slip plane did depend upon the location of the plane of maximum shear stress; for example, that if the plane of maximum resolved shear stress fell between planes of the types {110} and {123}, those planes would slip. However, Gough did not consider that variations of the ratios of the critical shear stresses could change the behavior of the material. The state of stress he investigated, alternating torsion, is different from that discussed in the following treatment, which is simple tension.

The basic notions and quantities used in the derivation of the resolved shear stress equations are given in Fig. 2. The direction of the tensile force F and the normal to the slip plane are denoted by the indices

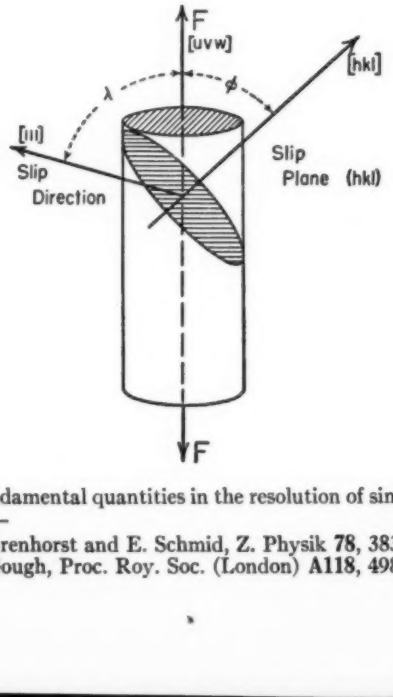


FIG. 2. Fundamental quantities in the resolution of simple tension.

* Now associated with Research Laboratories Division, General Motors Corporation. Submitted in partial fulfillment of the requirements for the degree of Doctor of Science at Carnegie Institute of Technology.

[1] A. Opinsky and R. Smoluchowski, Phys. Rev. **74**, 343 (1948), and R. Smoluchowski, Physica **15**, 179 (1949).

[2] E. N. da C. Andrade, Proc. Phys. Soc. (London) **52**, 1 (1940).

[3] Barrett, Ansel, and Mehl, Trans. Am. Soc. Metals **25**, 702 (1937).

[4] W. Fahrenhorst and E. Schmid, Z. Physik **78**, 383 (1932).

[5] H. J. Gough, Proc. Roy. Soc. (London) **A118**, 498 (1928).

$[uvw]$ and $[hkl]$, respectively. The slip direction, according to almost all investigations, seems to be $\langle 111\rangle$ and is affected neither by composition nor by temperature. On the other hand, the active slip planes do depend upon these two factors and can be any planes of the types $\{110\}$, $\{112\}$, or $\{123\}$. Of course, the slip plane has to belong to the zone of the slip direction. For instance, the zonal relation for the $[111]$ direction is $h+k+l=0$. If A denotes the cross-sectional area of the sample perpendicular to the direction of the tensile force, then the cross-sectional area of the plane of slip is given by $A/\cos\phi$, where ϕ is the angle between the normal to the slip plane and the tensile axis. The force resolved in the $[111]$ direction is $F\cos\lambda$, where λ is the angle between $[111]$ direction and the tensile axis, and so the resolved shear stress on the slip plane along the $[111]$ direction is

$$\tau^{111}=\frac{F}{A}\cos\phi\cos\lambda \tag{1}$$

which, by expressing the cosines in terms of Miller indices, can be written as

$$\tau^{111}=\frac{F}{\sqrt{3}A}\frac{(u+v+w)(hu+kv+lw)}{(h^2+k^2+l^2)^{\frac{1}{2}}(u^2+v^2+w^2)}. \tag{2}$$

For the slip direction $[\bar{1}11]$, one obtains

$$\tau^{\bar{1}11}=\frac{F}{\sqrt{3}A}\frac{(-u+v+w)(hu+kv+lw)}{(h^2+k^2+l^2)^{\frac{1}{2}}(u^2+v^2+w^2)} \tag{3}$$

and similar expressions for $\tau^{1\bar{1}1}$ and $\tau^{11\bar{1}}$. Maximum resolved shear stress occurs for $\lambda=\phi=45°$ as can be verified by differentiating (1). This means that on a stereographic projection the tensile axis must lie on the great circle joining the pole of the slip plane and the slip direction, and it must bisect the angle between them. For each permissible slip plane and slip direction combination, an orientation of the tensile axis can be found which corresponds to maximum resolved shear stress. For convenience, all subsequent consideration will be referred to the standard unit stereographic triangle formed by the poles (001), (011), and $(\bar{1}11)$ (see Fig. 3, which was taken from Barrett)[6] within which any tensile direction can be represented by a point.

In Fig. 4, the maxima for the various plane and direction combinations in or near the unit stereographic triangle are shown. These plane and direction combinations are the only ones which have maximum resolved shear stress in or near the unit stereographic triangle, and they are the only ones which need be considered. This result can be obtained in the following way. Using the interrelating equation $h+k+l=0$, and partially differentiating (2), with respect to h and k, one has the

[6] C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 34.

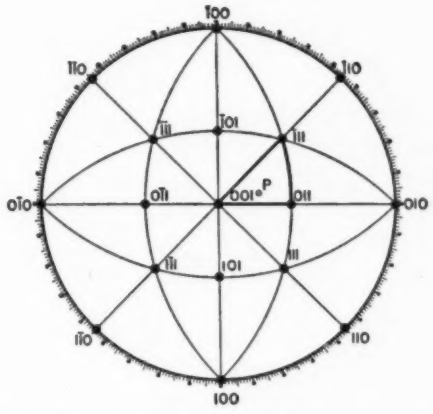


FIG. 3. The unit stereographic triangle in a (001) projection of a cubic lattice.

indicial equations

$$k=-h\frac{u-2v+w}{2u-v-w}, \tag{4}$$

$$l=-h-k=h\left(\frac{u-2v+w}{2u-v-w}-1\right) \tag{5}$$

and

$$|\tau_{\max}{}^{111}|=\frac{\sqrt{2}F}{3A}\frac{(u+v+w)}{(u^2+v^2+w^2)}\times(u^2+v^2+w^2-uv-uw-vw)^{\frac{1}{2}}. \tag{6}$$

For other slip directions equations similar to (6) can be obtained. As indicated in Fig. 4 for $[uvw]$ contained within the unit stereographic triangle, the $[111]$ slip direction has maximum resolved shear stress in one part of the triangle, while the $[\bar{1}11]$ direction has maximum resolved shear stress in another part.

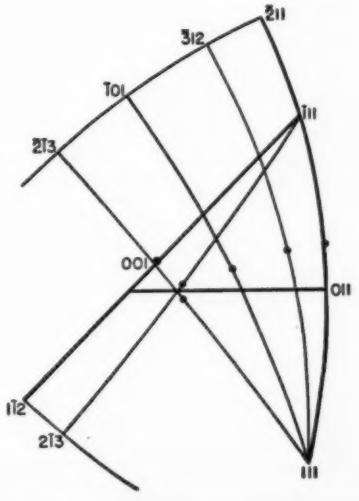


FIG. 4. Maxima of resolved shear stresses for various slip systems in the body-centered cubic lattice in the proximity of the unit triangle.

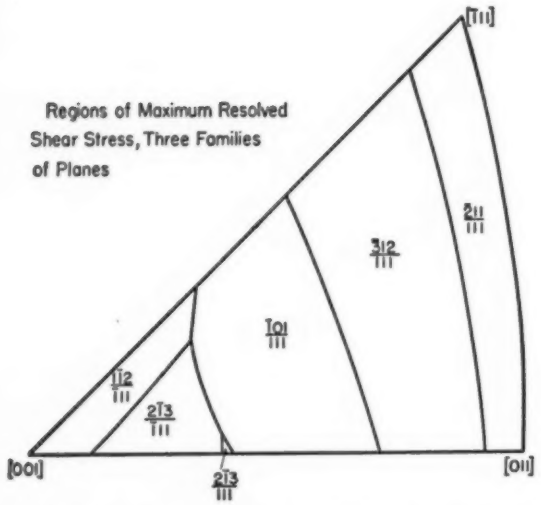


FIG. 5. Regions in a unit triangle within which the indicated plane and direction combination has resolved shear stress greater than any other system: three families of planes.

The next problem is to find, within the unit stereographic triangle, the boundaries between regions in which one of such plane-direction combinations has a greater value of the resolved shear stress than all the others. These boundaries can be calculated by substituting the values of h, k, and l into Eqs. (2) and (3), equating, and solving for u, v, and w. The result is shown in Fig. 5. Here the various regions of maximum resolved shear stress, for three families of planes, are shown as a function of the orientation of the tensile axis. It must be emphasized that Fig. 5 refers to the slip behavior only when the critical shear stresses are the same on the three types of planes. If the critical shear stresses are not all equal, the boundaries shown in Fig. 5 are shifted accordingly. If more than three types of planes are active, then there are additional regions in the unit

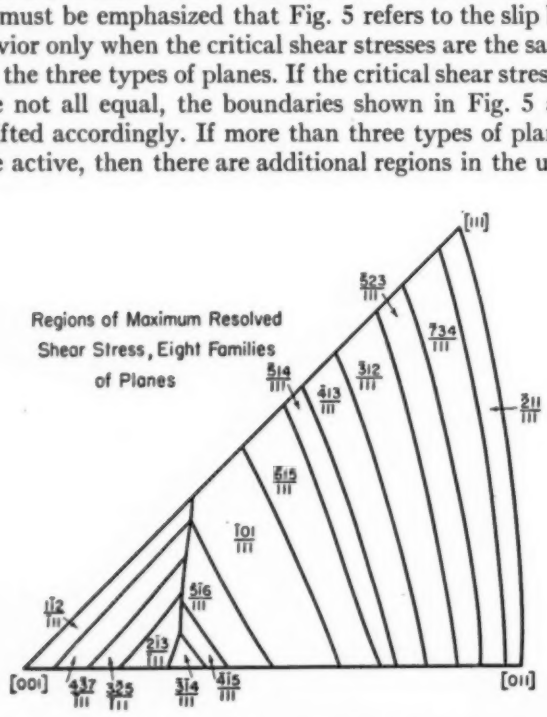


FIG. 6. Regions in a unit triangle within which the indicated plane and direction combination has resolved shear stress greater than any other system: eight families of planes.

stereographic triangle as shown in Fig. 6 for eight families of planes.

The dependence of the resolved shear stress for the various slip planes on the direction of the tensile axis along the [001]–[011] line of Fig. 5 is shown in Fig. 7. The intersections correspond to the boundaries in Fig. 5, which were obtained analytically rather than by the slow process of making sections through the unit stereographic triangle.

Equations (4) and (5) show that a given plane may have maxima in resolved shear stress for various values of u, v, and w. It appears that a plane (hkl) has maximum resolved shear stress for a particular $\langle 111 \rangle$ direction along a great circle passing between the normal to the slip plane and the slip direction. This is a more general statement of the conditions illustrated in Fig. 5. For the [111] direction, $(\bar{1}01)$ has maximum resolved shear stress all along the line from $[\bar{1}01]$ to [111]; however, maxima along other directions may exceed the value of the maximum on $(\bar{1}01)$ outside of the unit stereographic

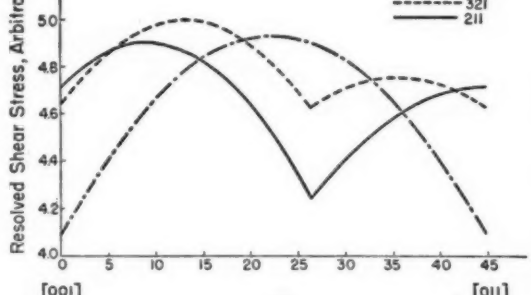


FIG. 7. Dependence of the resolved shear stress for various slip planes as a function of the orientation of the tensile axis along the line [001]–[011].

triangle considered. The maximum resolved shear stress of $0.5F/A$ is merely a special instance of the maximum, but it does fall on the great circle between the pole of the plane and the slip direction. Since the plane of maximum resolved shear stress remains constant along such a great circle, it is then a corollary that the ratios of the resolved shear stresses remain constant along the great circle. This can be seen intuitively from Fig. 8, which shows the shear stress resolved along [111] as a function of the planes of the zone [111] for various tensile axes along the line between [011] and $[\bar{1}11]$. For each direction, the maximum resolved shear stress falls on $(\bar{2}11)$; this is to be expected because the line [011]–$[\bar{1}11]$ is part of the great circle between [111] and the pole of $(\bar{2}11)$. To be sure, the maxima have different values at the various tensile axes, but nevertheless they still occur on $(\bar{2}11)$. The curves of Fig. 8 are sine curves because in Eq. (1) λ is held constant while ϕ is varied. It can be seen that along the line of tensile axes which passes between the pole of $(\bar{2}11)$ and the slip direction,

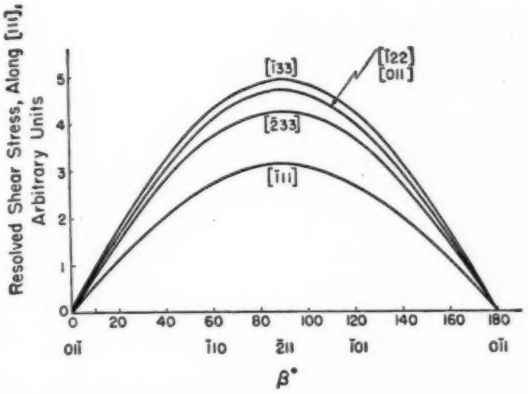


FIG. 8. Shear stress resolved along [111] for various planes in the [111] zone as a function of the orientation of the tensile axis along the line [111]–[2̄11].

the ratios of the shear stresses on the various planes remain constant. For example, the ratios of the shear stresses resolved on (2̄11) to those on (1̄10) and (3̄21) remain constant at each tensile axis shown on Fig. 8.

There are two important consequences of this behavior. The first is that slip, once it is initiated, will continue on the same plane until a boundary with a region belonging to another slip system is reached. Various investigations have shown that the tensile axis tends to rotate toward the slip direction during tensile elongation. Hence the tensile axis is moving along a great circle passing through the slip direction and the original tensile axis. Therefore, the ratios of the resolved shear stresses on the various possible slip planes belonging to the original slip direction remain constant. The second consequence of this property of the great circle is the fact that it is possible to project all data obtained in the unit stereographic triangle on a line, and thus one can reduce the complexity of the problem by one dimension. This implies also that fewer experimental points are required in the triangle to obtain significant conclusions.

2. THE ORIENTATION DEPENDENCE OF SLIP BEHAVIOR

The previous section dealt with the orientation dependence of the resolved shear stress; here the orientation dependence of the resulting slip will be discussed.

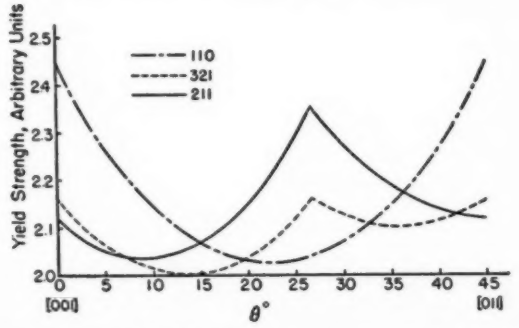


FIG. 9. Yield strength (arbitrary units) as a function of the orientation of the tensile axis along the line [001]–[011].

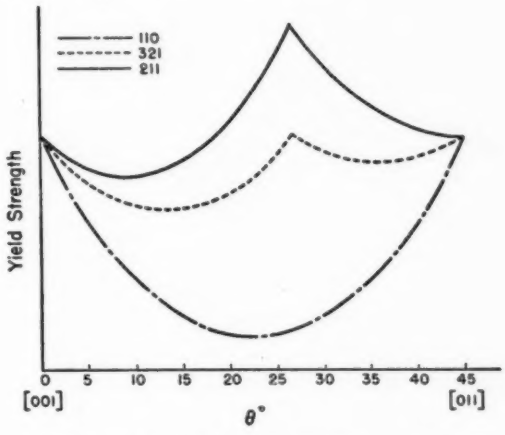


FIG. 10. Limiting case for {110} slip alone.

According to the critical shear stress law, slip occurs when the resolved shear stress on the plane along the slip direction, τ, is equal to the critical shear stress, S. Since the yield strength $Y=F/A$, we have

$$Y=S/\cos\lambda\cos\phi. \tag{7}$$

Thus, if all critical shear stresses are equal, the plot of yield strength *versus* orientation of the tensile axis within the unit stereographic triangle will be identical with Fig. 6. If the critical shear stresses are not all equal, then the yield strength curves and their intersections will be shifted. This is illustrated in Fig. 9, which shows the yield strength for the orientations along the line [001]–[011], for all three types of planes having equal critical shear stresses. Figure 9 is merely the inverse of Fig. 7 as indicated by Eq. (7). If the critical shear stresses are not equal, the curves in Fig. 9 will be displaced vertically with respect to one another with a minor change in shape which can be neglected as shown below.

If the Y_{110} curve is displaced below the other two curves, Fig. 10, slip will occur only on {110} planes for all orientations of tensile axes. One can calculate the ratios of the critical shear stresses necessary to produce this radical change: For slip to occur on {110} alone, one need set the yield strength curves to coincide at (001) and (011) and, knowing the resolved shear stress for the various types of planes at these orientations, calculate

TABLE I. Behavior of the yield strength curves for various ratios of the critical shear stresses.

Value of ratio	S_{112}/S_{110}	S_{123}/S_{110}	S_{123}/S_{112}
$0 < r < 0.866$	$Y_{112} < Y_{110}$	$Y_{123} < Y_{110}$	$Y_{123} < Y_{112}$
$0.866 \leqslant r < 0.945$	$Y_{112}\ i\ Y_{110}$	$Y_{123} < Y_{110}$	$Y_{123} < Y_{112}$
$0.945 \leqslant r < 0.982$	$Y_{112}\ i\ Y_{110}$	$Y_{123}\ i\ Y_{110}$	$Y_{123} < Y_{112}$
$0.982 \leqslant r \leqslant 1.091$	$Y_{112}\ i\ Y_{110}$	$Y_{123}\ i\ Y_{110}$	$Y_{123}\ i\ Y_{112}$
$1.091 < r \leqslant 1.134$	$Y_{112}\ i\ Y_{110}$	$Y_{123}\ i\ Y_{110}$	$Y_{123} > Y_{112}$
$1.134 < r \leqslant 1.155$	$Y_{112}\ i\ Y_{110}$	$Y_{123} > Y_{110}$	$Y_{123} > Y_{112}$
$1.155 < r$	$Y_{112} > Y_{110}$	$Y_{123} > Y_{110}$	$Y_{123} > Y_{112}$

The designation i stands for intersects.

TABLE II. Ratios of the critical shear stresses for various tensile axes along the line [001]–[011].

Tensile axis	Angle from (001)		S_{110}/S_{112}	S_{123}/S_{112}	S_{123}/S_{110}
001	0°	0′	0.866	0.982	1.133
01 10	5	43	0.912	0.999	1.096
016	9	28	0.945	1.012	1.071
014	14	2	0.990	1.029	1.039
013	18	26	1.039	1.048	1.008
025	21	48	1.083	1.064	0.983
012	26	34	1.155	1.091	0.945
035	30	58	1.083	1.064	0.981
023	33	41	1.039	1.047	1.005
034	36	52	0.990	1.029	1.042
078	41	12	0.924	1.004	1.084
011	45		0.866	0.982	1.134

the ratio of the critical shears. Thus, from (9) we have $Y_{110} \leqslant Y_{123}$, $(S_{110}/0.408) \leqslant (S_{123}/0.462)$ or $(S_{123}/S_{110}) \geqslant 1.13$. In a similar manner, the ratios of the critical shear stresses on {110} and {123} can be calculated to permit slip on {110} alone, and the resulting relation is $(S_{112}/S_{110}) \geqslant 1.16$. Other such calculations can be made; they are summarized in Table I.

An example will clarify the use of the table. Let $S_{110}=1000$, $S_{112}=1200$, and $S_{123}=1200$. Thus $(S_{112}/S_{110})=1.2$, $(S_{123}/S_{110})=1.2$, and $(S_{123}/S_{112})=1$. Therefore, respectively, $Y_{112}>Y_{110}$, $Y_{123}>Y_{110}$, and Y_{123} intersects Y_{112}. However, since Y_{110} lies below both Y_{123} and Y_{112}, slip will occur on {110} alone. Table I shows also that significant changes in the choice of the operative slip plane occur only when the critical shear stresses differ by less than 16 percent. These small differences produce negligible changes in the shapes of the curves for individual slip planes in Fig. 9 while the relative position of the curves can be appreciably altered.

If the critical shear stresses or their ratios are known, Table I will give information concerning the active slip planes of the material. However, usually the situation is reversed: the ratios of the critical shear stresses, which could produce the observed slip behavior, are to be determined from the known operative slip planes. Experimentally, with the aid of the projection properties of great circles passing through the slip direction, the boundaries between operative slip planes along the line of tensile axes [001]–[011] can be determined. This could be done either by determination of the yield strength values, with a subsequent calculation of projected yield strength values along the base line, or by observation of the operative slip planes by metallographic means.

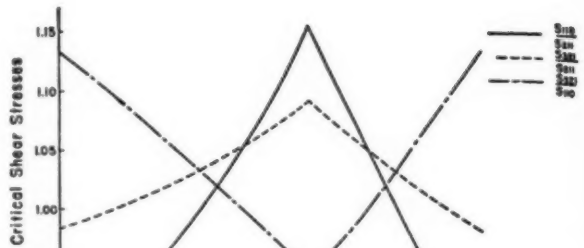


FIG. 11. Ratios of the critical shear stresses as a function of the orientation of the tensile axis along the line [001]–[011].

The choice of the line [001]–[011] as the base line for such a projection is particularly advantageous. It can be shown that, the resolved shear stresses for the same type of plane, having [111] and [$\bar{1}$11] as their slip directions, coincide at each tensile axis along this line. Putting $u=0$, which corresponds to the line (001)–(011), into (2) and (3) shows that then $\tau^{111}=\tau^{\bar{1}11}$. Thus, only one set of master curves need be plotted to permit comparison of experimental data with calculations, although two different slip directions may have been operating. The ratios of the critical shear stresses as a function of the orientation of the tensile axis along the base line [001]–[011] can be easily calculated from Eq. (7) and are given in Table II, and plotted in Fig. 11.

The use of these results is very simple: for instance, if the boundary between operative slip planes of the types {110} and {123} occurs at 38°, then the ratio of the critical shear stresses is approximately 1.05. Application of this method to various experimental data will be give in a subsequent paper.

This work has been sponsored by the AEC contract. One of the authors (A.J.O.) wishes to express his appreciation to the Atomic Energy Commission for a fellowship.

Letters to the Editor

Equivalent Temperature of an Electron Beam

M. E. HINES
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received July 13, 1951)

1. Introduction.—Parzen and Goldstein[1] in a recent article have shown the effects of a distribution of electron velocities on the gain of a traveling wave tube. For their assumption that the longitudinal velocity distribution has an equivalent temperature equal to that of the cathode, substantial reduction in gain is predicted for some types of tubes.

The purpose of this letter is to discuss the equivalent temperature to be expected in the beam of a practical traveling wave tube. It is shown that the effect of beam acceleration is to reduce the spread in longitudinal velocities by a large factor for cases of nonconverging electron guns. The effects of long drift times and converging guns are discussed in a qualitative manner. These effects can result in a high beam temperature from other considerations.

2. Temperature Calculation.—For this derivation it is assumed that the beam is accelerated from the cathode in substantially rectilinear flow with the applied fields acting in the longitudinal direction only. It is also assumed that the electrons do not interact with each other on a microscopic scale.

The distribution function of the longitudinal velocities at the potential minimum just in front of the cathode is

$$dn=(e\bar{n}/kT_c)\exp(-eV^*/kT_c)dV^*,$$

where $\bar{n}$ is number passing across per unit time and

$$V^*=u_{zm}^2/(2e/m).$$

Here u_{zm} is the electron velocity at the potential minimum. After acceleration to a potential V_0 above cathode potential, an electron with an initial kinetic energy eV^* at the minimum will have a final velocity given by

$$u_z=(2(e/m)(V_0+V_m+V^*))^{\frac{1}{2}},$$

where $-V_m$ is the potential at the minimum. If V^* is very small in comparison to V_0, the velocity may be closely approximated by

$$u_z=(2(e/m)(V_0+V_m))^{\frac{1}{2}}[1+(V^*/2(V_0+V_m))].$$

The mean velocity of the stream is given by

$$\bar{u}_z=\frac{1}{n}\int_n u\,dn \quad \text{(see footnote *)}$$

$$=\frac{e}{kT_c}\int_0^\infty\left(2\frac{e}{m}(V_0+V_m)\right)^{\frac{1}{2}}\left[1+\frac{V^*}{2(V_0+V_m)}\right]\exp\left(-\frac{eV^*}{kT_c}\right)dV^*$$

$$=(2(e/m)(V_0+V_m))^{\frac{1}{2}}[1+(kT_c/2e(V_0+V_m))].$$

The equivalent beam temperature is given by

$$T_b=(1/kn)\int m(u_z-\bar{u}_z)^2dn$$

$$=\frac{e^2}{2k^2T_c(V_0+V_m)}\int_0^\infty\left[V^*-\frac{kT_c}{e}\right]^2\exp\left(-\frac{eV^*}{kT_c}\right)dV^*.$$

$$T_b=\left[\frac{kT_c}{2e(V_0+V_m)}\right]T_c.$$

3. Effects of Long Drift Times.—For $V_0=1000V$ and $T_c=1000°K$, the above formula predicts an effective beam temperature of about 1/20 of a degree absolute, rather than 1000°. This is the temperature corresponding to the longitudinal velocity spread only. The transverse velocity spread remains the same as it was at the time of emission, which corresponds to the temperature of the cathode.

If the beam is allowed to drift for a sufficiently long time, these transverse velocities will have their energies redistributed so that each degree of freedom has an equal thermal velocity distribution. This would imply that the final beam temperature should approach approximately ⅔ of the cathode temperature if allowed to drift for an indefinite distance. There is, however, some evidence that the microscopic interactions between the electrons are sufficiently infrequent or ineffective that it may take a very long drift time to re-establish equilibrium in practical tubes, depending upon the charge density in the beam.

One such piece of evidence can be deduced from the measurements of the noise currents in an electron beam reported by Cutler and Quate.[2] In the theories of Pierce,[3] Rack,[4] and Peterson,[5] the velocity fluctuations at the cathode are assumed to be the major source of noise in vacuum tubes. These velocity fluctuations cause fluctuations in electron transit times during acceleration and drift which result in convection current and velocity fluctuations which are distinctly correlated with one another. In a drifting beam, such correlated fluctuations will cause a repeating pattern of interference between the space-charge waves so established. Cutler and Quate[2] observed these interference effects and they show that they are readily explainable in terms of the theories of Rack, Pierce, and Peterson. The waves they observe are of such a magnitude that the velocity fluctuations at the maximum are rather close to that which Rack gives for the fluctuations at the cathode, namely

$$v^2=\frac{(4-\pi)(e/m)kT_cB}{I_0}.$$

The fact that such maximum fluctuations were observed does not imply that the thermal velocity distribution in the final beam corresponds to the cathode temperature as these fluctuations are produced by a wave propagation phenomenon which involves the beam as a whole. Rather, one would expect a considerably greater magnitude of noise if the beam had acquired an equilibrium maxwellian velocity distribution by interaction with the transverse thermal velocities. Rack's formula applies to the unsymmetrical velocity distribution at the potential minimum next to the cathode surface. The appropriate formula for the velocity fluctuations in a symmetrical random maxwellian velocity distribution is

$$v^2=\frac{4(e/m)kTB}{I_0},$$

which is about 5 db larger for an assumed equilibrium temperature ⅔ that of the cathode. If thermal equilibrium had been established then it might be expected that new velocity fluctuations of this larger magnitude would have appeared in the process, generating space charge waves in the beam with associated noise currents of larger magnitude. Cutler and Quate did not observe such effects so that it might be deduced that thermal equilibrium had not been established in their beam. This would imply that the effective beam temperature, considering the longitudinal velocity distribution only, was much lower than the cathode temperature in their experiment.

4. Effects of Beam Focusing.—A sharply converging electron gun can result in a rather wide distribution of longitudinal velocities. This will occur because of the transverse electron velocity distribution in the accelerating and focusing region. The beam contains many electrons which have not followed the ideal trajectories and consequently have been accelerated in an abnormal manner so that they have rather widely dispersed transverse velocities. These electrons must also have a dispersion in their longitudinal velocities because of energy considerations. This can increase the effective beam temperature markedly for guns with a high degree of beam convergence.

Methods of beam focusing which cause a variation in dc velocity over the beam cross section will cause effects similar to those of high beam temperature. If the beam is focused in the ideal "Brillouin flow" condition,[6] no difficulty of this sort is to be

expected as the longitudinal velocity of the electrons is uniform throughout the beam.

5. Summary.—The equivalent temperature of the longitudinal velocity distribution in an electron beam may be very low in cases where the electron gun is not of the converging type and where the beam is focused in the "Brillouin flow" condition. If a beam is allowed to drift for an indefinite time, then the transverse velocities may re-establish an equilibrium at about $\frac{2}{3}$ the cathode temperature, but there is evidence that this may take a rather long drift time. Converging electron guns and improper focusing conditions will increase the effective beam temperature.

[1] Philip Parzen and Ladislas Goldstein, J. Appl. Phys. **22**, 398–401 (1951).
[2] Strictly speaking, dn in this formula specifies the distribution function per unit volume, rather than the number passing per unit time. However, because of the small velocity spread, the difference between the two forms is negligible in this case.
[3] C. C. Cutler and C. F. Quate, Phys. Rev. **80**, 875 (1950).
[4] J. R. Pierce, *Traveling Wave Tubes* (D. Van Nostrand Company, Inc., New York, 1950), pp. 148–155.
[5] A. J. Rack, Bell System Tech. J. **17**, 592 (1938).
[6] L. C. Peterson, Proc. Inst. Radio Engrs. **35**, 1264 (1947).
[7] J. R. Pierce, *Theory and Design of Electron Beams* (D. Van Nostrand Company, Inc., New York, 1950), pp. 152–155.

Comments on "Equivalent Temperature of an Electron Beam"

PHILIP PARZEN
Federal Telecommunication Laboratories, Nutley, New Jersey
(Received August 10, 1951)

MR. HINES'S derivation of the variation of temperature in a rectilinear beam is correct if the electron-electron interaction is neglected. A similar derivation may be found on page 126 of the book, *Klystrons and Microwaves Triodes*, by Hamilton, Knipp, and Kuper of the M.I.T. Series (McGraw-Hill Book Company, Inc., New York, 1948). Mr. Hines's result is most easily obtained from Eq. (40) of the mentioned paper by Parzen and Goldstein. Thus, assuming no interaction, $q=0$ and hence $Tv^2=$constant. However, the electron-electron interaction may not be neglected. The experiments of Culter and Quate have not as yet been sufficiently interpreted to indicate the degree of interaction. There are effects due to the finite lateral extension of the electron beam which remain to be explained. Accurate measurements of the period of the space charge waves in the drift tube may cast some light on the actual electron temperature in the drift tube.

A Note on the "Attenuation of Radio Signals Caused by Scattering"

J. B. SMYTH AND C. P. HUBBARD
U. S. Navy Electronics Laboratory, San Diego, California
(Received July 2, 1951)

IN a recent paper[1] to the Journal of Applied Physics, A. H. LaGrone, W. H. Benson, Jr., and A. W. Straiton have integrated the Booker Gordon equation to obtain an expression for the attenuation of a plane wave due to scattering. Referring to Eq. (14) of their paper, it is seen that the shorter wavelengths are more rapidly attenuated, and indeed if we consider a limiting case, light should scarcely be propagated at all! A glance at Figs. 2 and 3 of the paper[1] indicates an error, for the curves of attenuation have no maxima and increase indefinitely with increasing l/λ. In deriving the expression for attenuation, the authors[1] have subtracted the power scattered in all directions from the incident power density, including the power scattered in the forward direction. In considering a plane wave, however, power scattered in a direction having a forward component is a part of the propagated wave. It is suggested, therefore, that the integration of the Booker Gordon equation be performed, not over the entire surface S, but

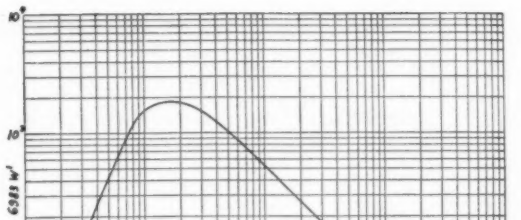


FIG. 1.

only over the hemisphere backward to the direction of propagation Denoting by W' the power scattered by unit scattering volume per unit incident power density in all directions making an angle equal to or greater than 90° with the direction of propagation, we have

$$W'=\int_{\theta=\pi/2}^{\pi}\int_{\delta=0}^{2\pi}\sigma(\theta,x)\sin\theta d\theta d\delta.$$

This integration yields

$$W'=\frac{\langle(\Delta\epsilon/\epsilon)^2\rangle\pi}{\lambda(2\pi l/\lambda)}\left\{\frac{1+5(2\pi l/\lambda)^2+8(2\pi l/\lambda)^4}{2+12(2\pi l/\lambda)^2+15(2\pi l/\lambda)^4}-\frac{1+2(2\pi l/\lambda)^2}{4(2\pi l/\lambda)^2}\log_e\frac{1+4(2\pi l/\lambda)^2}{1+2(2\pi l/\lambda)^2}\right\}.$$

Using W' as given by this expression, instead of the W given by Eq. (9), we arrive at more satisfactory attenuation curves. We submit, as a substitute for Fig. 2, a plot of the value of $6983W'$ vs the ratio of l/λ for the special case of $[\langle(\Delta\epsilon/\epsilon)^2\rangle/\lambda]=1$ (see Fig. 1). Unfortunately, neither the original approach[1] nor the one suggested in this note can be used to solve the problem indicated by the title of the paper.

[1] LaGrone, Benson, Jr., and Straiton, J. Appl. Phys. **22**, 672 (1951).

Reply to J. B. Smyth and C. P. Hubbard

A. H. LAGRONE
The University of Texas, Austin, Texas
(Received July 25, 1951)

DR. J. B. Smyth and Mr. C. P. Hubbard of the Navy Electronics Laboratory, San Diego, California, have proposed that the paper "Attenuation of radio signals caused by scattering"[1] be modified to the extent that only backscattering be considered as the cause of attenuation. They point out that all scattering at $\theta<90°$ has a forward component and should be considered as part of the propagated wave.

It is difficult for the authors to see where such a modification would be an improvement to the original paper, if indeed, the proposed modification has any merit at all. The original paper reported the power scattered per unit macroscopic element of volume relative to the power incident on the volume. It was recognized that a fraction of this power would be propagated in the same direction as that of the incident wave, and that for large values of (l/λ), a receiver along the path and in the immediate vicinity of the scattering would receive an appreciable part of the

total scattered power. At the same time, the effective area of a more distant receiver would be so small compared to the principal area over which scattering occurred that essentially all of the scattered power would be lost. Such considerations as these, however, constitute specific problems and require specific solutions.

The limiting case of light, as considered by Dr. Smyth and Mr. Hubbard, is for (l/λ) very large. This means that the scattered power is principally confined to a relatively small angle about $\theta=0°$. As long as (l/λ) is not infinite, however, this elementary solid angle is finite and the problem is similar to the one above. The original paper[1] was concerned with the scattering of radio waves and was not intended to include light waves.

It was not within the scope of the original paper[1] for the authors to consider every physically realizable field problem, rather, it was the aim to show the characteristics of the total power scattered in terms of the propagation parameters. The proposal of Dr. Smyth and Mr. Hubbard does not in any way modify these facts.

[1] LaGrone, Benson, Jr., and Straiton, J. Appl. Phys. **22**, 672–674 (1951).

Distribution of Slip in Metal Crystals

F. C. Frank*
The H. H. Wills Physical Laboratory, The University, Bristol, England
(Received July 30, 1951)

Kurnosov, Tronina, and Yakutovich[1] consider that their measurements show that the major part of the deformation in zinc crystals (compressed a few percent) is homogeneous, not localized in slip packets. Their data can be used to derive the opposite conclusion.

They performed experiments on zinc single crystals at 200°C, 19°C, and −190°C. They present no results for the first and no details for the last. Let us therefore concentrate on the room temperature study. Their result, that only 9 percent of the slip is in slip packets, was deduced from measurements on a specimen compressed 4 percent. Details and photographs are not given for this experiment. The details that are given refer to studies of slip band size and distribution at five different compressions up to 1.5 percent. One has the impression that already at 1.5 percent compression they may have missed some part of the "packet-slip" because of displacements by more than one fringe occurring within the lateral resolution of their instrument. They mention no test (such as change of wavelength) that might have detected such errors. The only wavelength stated to be used is 0.5×10^{-3} mm defined by a green filter. Their Fig. 4 shows several places where it would be desirable to test for whole-fringe displacements, including a place where the visible displacement is zero, but a dark line parallel to slip lines crosses the fringes.

A test discriminating between slip concentrated in slip packets, and homogeneous slip, is to find whether the rotation of the surface between slip packets is equal to the rotation of glide-planes, which can be calculated from the degree of compression if slip occurs on one family of planes only. It should be equal if slip occurs entirely in the packets, and less if some of the slip occurs homogeneously (or apparently homogeneously, within the resolution of the optical equipment used).

The profiles shown in their Fig. 5 allow one to distinguish sharply between slip-packet regions and interpacket regions. These are of about equal width at 1.5 percent deformation, while at 0.7 percent deformation the total of interpacket regions is twice as wide as the total of slip-packet regions. The inclination of the surface to the axis in slip-packet regions (average from Table IV) is 1° at 1.5 percent compression, 0.8° at 0.7 percent compression. If the definition of specimen axis is such that the mean-free-surface of the specimen does not rotate with respect to it (as is implied by the authors' use of Eq. (4)), it follows that the rotation in interpacket regions is about 1° for 1.5 percent, 0.4° for 0.7 percent compression. The lattice plane rotations calculated from change of length are 45′ and 21′, respectively. These figures are in approximate agreement and the discrepancy is in the opposite direction from that which would result from slip in the interpacket region. The inference is that within the uncertainty of the data, 100 percent of the slip is in the slip packets.

The sign of the discrepancy is nevertheless the contrary one to that which we should anticipate from our suspicion that some whole-fringe displacements have been missed, on the reasonable presumption that these errors mainly occur in the regions where some slip is recognized. A rather probable error is that in spite of using their Eq. (4),

$$\sin\chi_1=(L_0/L_1)\sin\chi_0,$$

which implies that the slant direction on the specimen surface is taken as axis once deformation occurs, the authors have nevertheless measured their angles from the normal to the pressure plates rather than from this slant axis. This supposition is supported by their Fig. 5, in which, however, no reference axis is specified. If this is the case they have wrongly added in a rotation,

$$\delta\psi=\sec\chi(\operatorname{cosec}\chi-2\sin\chi)\delta(\ln L).$$

(This formula applies for small compressions.) For their specimen (4), on which the reported measurements were made, χ was 40°, and $\delta\psi$ is therefore 25′ at 1.5 percent compression, 11′ at 0.7 percent compression. The mean measured inclinations of slipped regions of the surface to the surface to the specimen axis would then be 0.6° in each case; whence the rotation of unslipped regions would be 36′ and 18′, compared with the calculated values 45′ and 21′. The sign of the discrepancy is now that which we should expect if some part of the slip escaped notice. It also increases with increasing deformation, as we should expect.

The common hypothesis, accepted on inadequate evidence, is that most of the slip in metal crystals is associated with visible slip bands. It would not have been surprising if Kurnosov, Tronina, and Yakutovich, using a microinterferometer of 96× lateral magnification, had failed to see all the slip bands present, and so reached a contrary conclusion. While the quality of their observations is scarcely sufficient to make a useful contribution towards settling this question, they nevertheless support the usual hypothesis, rather than, as they say, contradict it.

* Temporarily at General Electric Research Laboratory, The Knolls, Schenectady, New York.
[1] Kurnosov, Tronina, and Yakutovich, J. Tech. Phys. (U.S.S.R.) **18**, 197 (1948).

Negative Charging of Glass Fibers under Electron Bombardment

John H. L. Watson
Edsel B. Ford Institute for Medical Research, Detroit, Michigan
(Received August 13, 1951)

In a paper by S. F. Kern and R. A. Kern[1] an explanation based upon the existence of a charge differential between the object (in their case, round particles from Dow 580 G latex) and its surrounding substrate is advanced for part of the increased image dimensions of shadow cast over nonshadow cast objects in electron microscopes. Their argument being, that if sufficient charge differential could be produced and the object was negatively charged, the negative lens so formed would lower the magnification, giving smaller and erroneous dimensions in unshadowed fields.

The presence of a charge has not been proven for the Dow latex, and in any case, the likelihood of positive rather than negative charges on the object would have to be considered too. However, it is possible to show that some nonconducting objects of macroscopic dimensions can become charged under electron bombardment in electron microscopes and that these charges can be negative. The following observations are not offered as a general explanation of the phenomenon of enlarged image dimensions in shadowed fields but as one example of the definite existence of a negative charge existing on one type of specimen.

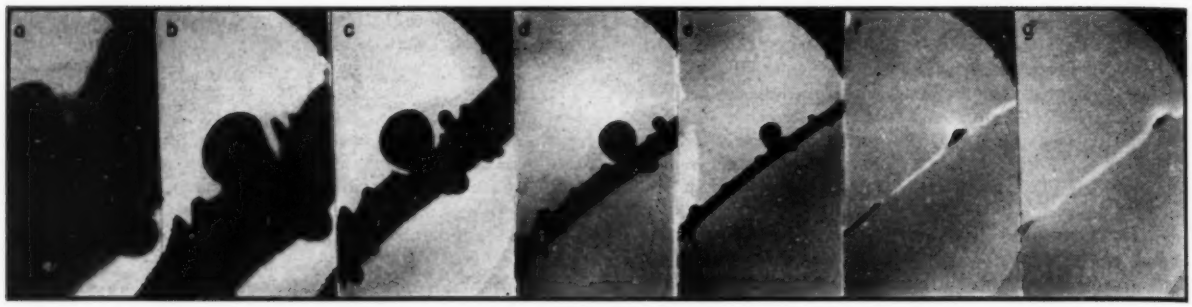


FIG. 1.

The specimen chosen is a glass fiber of one micron diameter fastened firmly with rubber cement across the open end of an electron microscope specimen holder. The cap is put on without a specimen screen.

In Fig. 1 a series of electron micrographs[2] are shown of the glass fiber. Because of its high curvature an appreciable electrostatic field is built up about it when it is within the electron beam. It apparently contracts a negative charge and acting as a simple, negative lens diverges the beam to raise the apparent position of the object and lower the magnification of the image with normal focusing current through the objective coil.

Figure 1 illustrates the appearance of the image when the objective lens current is varied from a lower-than-focus to a higher-than-focus value, with the beam intensity held fairly high and constant. Under these conditions it is possible to change the image from one which is large and out of focus in (a), to one which is well focused and smaller in (d) and (e), and finally to one in which there is a condition analogous to dark field in (f) and (g). The image remains in fairly good focus over a range of lens current, but its size varies considerably over the same range, so that any reliable determination of fiber diameter is impossible unless the intensity is much reduced.

A similar series of micrographs may be obtained if the objective current is kept constant at focus, and the intensity varied from low to high except that the large, diffuse images of (a), (b), and (c) in the figure are not met. As the intensity is raised the focused image decreases in diameter and eventually the dark field condition recurs.

The microscopic particles clinging to the fiber are of methylene blue collected on the fiber and are not significant except to show that under conditions of mounting they too vary in dimensions with the fiber and are therefore charged similarly.

[1] S. F. Kern and R. A. Kern, J. Appl. Phys. 21, 705 (1950).
[2] Micrographs reproduced from Ph.D. thesis of John H. L. Watson, University of Toronto, 1943.

A Note on the Partial Differential Equations Describing Steady Current Flow in Intrinsic Semiconductors

R. C. PRIM
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received August 7, 1951)

THE steady flow of holes and electrons in the interior of a homogeneous intrinsic semiconductor subject to the assumption of uniform temperature, absence of space charge effects, and negligibility of internal recombination is described[1,2] by the following relations:

$$i = i_p + i_n$$

with

$$i_p = -\mu_p e\left(p\nabla V + \frac{kT}{e}\nabla p\right)$$

$$i_n = -\mu_n e\left(n\nabla V - \frac{kT}{e}\nabla n\right)$$

subject to

$$\nabla\cdot i_p = \nabla\cdot i_n = 0$$

and

$$n - p = 0, \quad p \geq 0.$$

[i: total current density vector; i_p: hole current density vector; i_n: electron current density vector; μ_p: hole mobility constant; μ_n: electron mobility constant; e: magnitude of electronic charge; k: Boltzmann's constant; T: absolute temperature; p: specific concentration of holes; n: specific concentration of electrons.]

Through consideration of the vectors

$$i_p/\mu_p e \mp i_n/\mu_n e,$$

these relations can be reduced[1] to the following pair of partial differential equations:

$$\nabla^2 p = 0,$$
$$\nabla\cdot(p\nabla V) = 0. \qquad [A]$$

The purpose of this note is to present the following

Theorem:

Let $V[r, z]$ be any axially symmetric harmonic function in the ordinary cylindrical coordinates (r, θ, z), and let $p[x, y]$, $q[x, y]$ be any conjugate pair of plane harmonic functions. Then the function pair

$$p[x, y]; \quad V[p(x, y), q(x, y)]$$

satisfies [A], and any plane solution of [A] for which $\nabla p \not\equiv 0$ can be so represented.

For the proof of this theorem, the conjugate harmonic functions p and q are introduced as independent variables in the second equation of [A]. Then

$$p\nabla V = p\frac{\partial V}{\partial p}\nabla p + p\frac{\partial V}{\partial q}\nabla q,$$

whence

$$\nabla\cdot(p\nabla V) = \frac{\partial}{\partial p}\left(p\frac{\partial V}{\partial p}\right)(\nabla p)^2 + \frac{\partial}{\partial q}\left(p\frac{\partial V}{\partial p}\right)(\nabla q\cdot\nabla p) + p\frac{\partial V}{\partial p}\nabla^2 p$$
$$+ \frac{\partial}{\partial p}\left(p\frac{\partial V}{\partial q}\right)(\nabla p\cdot\nabla q) + \frac{\partial}{\partial q}\left(p\frac{\partial V}{\partial q}\right)(\nabla q)^2 + p\frac{\partial V}{\partial q}\nabla^2 q.$$

But

$$\nabla^2 p = \nabla^2 q = \nabla p\cdot\nabla q = \nabla q\cdot\nabla p = 0,$$

and

$$(\nabla q)^2 = (\nabla p)^2;$$

hence

$$\left[\frac{\partial}{\partial p}\left(p\frac{\partial V}{\partial p}\right) + \frac{\partial}{\partial q}\left(p\frac{\partial V}{\partial q}\right)\right](\nabla p)^2 = 0.$$

Inasmuch as the equation satisfied by axially symmetric harmonic functions in ordinary cylindrical coordinates (r, θ, z) is

$$\frac{\partial}{\partial r}\left(r\frac{\partial \varphi}{\partial r}\right) + \frac{\partial}{\partial z}\left(r\frac{\partial \varphi}{\partial z}\right) = 0,$$

the theorem follows.

The special cases $\partial V/\partial p=0$ and $\partial V/\partial q=0$ are included in more general results in reference 1.

[1] R. C. Prim, Bell System Tech. J. **30**, 4 (1951).
[2] W. van Roosbroeck, Bell System Tech. J. **29**, 4, 560–607 (1950).

Noble Gas Hot Cathode Diodes with Negative Arc Drop

G. Medicus and G. Wehner
WCESV, Wright Field, Dayton, Ohio
(Received August 16, 1951)

PREVIOUS investigations of the low voltage arc in noble gas hot cathode diodes[1] had shown the decisive importance of the work functions of the electrodes, especially of the anode for the arc drop (which in principle was well known, for instance, from the early work on the retarding field current in vacuum tubes). These investigations led to the conception that by lowering the work function of the anode, and at the same time increasing the work function of the cathode negative, arc drops could be obtained. Therefore a diode with an indirectly heated Ta cathode was built, in order to provide an equipotential, high work function cathode. When the anode of this tube was covered with Ba, evaporated from a Ni-Ba wire, in order to obtain a low work function anode, negative voltage drops up to 0.8 v at amperages in the order of 0.5A were observed, without any detectable oscillations, at pressures in the order of 0.1 mm Xe. In this case the tube worked as an emf, where part of the heating energy of the cathode (about 0.3 percent) was directly transformed into electrical energy. As probe measurements showed, the discharge mechanism in this case in principle is the same as in the case of the positive arc drop previously investigated especially by Compton and Eckart[2] and by Druyvesteyn,[3] only the current-voltage characteristic is shifted so that a region of negative voltage drop appears. The basic fact which leads to the phenomenon described above is the external electrostatic field existing between the surfaces of any two metals in contact with each other, if they have different work functions. The electrons emitted by the high work function cathode roughly have a potential energy with respect to the anode equal to the difference in work function of cathode and anode which enables them to establish an emf.

A detailed report on this subject will be published later.

[1] G. Medicus and G. Wehner, Phys. Rev. **82**, 570 (1951).
[2] K. T. Compton and C. Eckart, Phys. Rev. **25**, 139 (1925).
[3] M. J. Druyvesteyn, Z. Physik **65**, 781 (1930).

The Temperature Function in a Moving Medium

R. H. Ritchie
Health Physics Division, Oak Ridge, Tennessee
(Received August 25, 1951)

A PAPER by B. Podolsky (J. Appl. Phys. **22**, 581 (1951)) treats the problem of the temperature distribution in an infinite homogeneous medium moving with constant velocity v with respect to a stationary point source of heat. The steady-state solution is found by a rather unwieldy method.

This solution, as well as the corresponding transient solution, may be obtained by the method of superposition of sources.[1] This method is not only more elegant and direct but lends itself to the solution of a wide variety of problems of this general nature.

The temperature at a point P with position vector r resulting from a stationary point source of heat which is liberated instantaneously at the origin in a stationary infinite medium is given by[2]

$$\frac{W}{8k\kappa^{\frac{1}{2}}(\pi t)^{\frac{1}{2}}}\exp\{-|r|^2/4\kappa t\},$$

where W is the quantity of heat liberated, t is the time measured from the instant of liberation, $\kappa(=k/c\rho)$ is the thermal diffusivity of the medium, k, c, and ρ are the thermal conductivity, specific heat, and density of the medium, respectively. The medium is assumed to be initially at zero temperature. By superposing instantaneous point sources and taking into account the motion of the medium past the point in question, one finds that the temperature at r due to a stationary volume source, $S(r)$, whose strength varies in time as $W(t)$ ($W(t)=0$ for $t<0$) is given by

$$T=\frac{1}{8k\kappa^{\frac{1}{2}}\pi^{\frac{1}{2}}}\int_0^t d\tau\int dr' S(r')W(t-\tau)\times\exp\{-|r'+r^*(t)-r|^2/4\kappa t\}\frac{1}{\tau^{\frac{1}{2}}},\quad(2)$$

where $r^*(t)$ is the position vector of the origin of a set of coordinate axes which moves with the medium and which coincides with the fixed system at $t=0$.

For the case treated by Podolsky, $S(r)=\delta(r)$ (the Dirac delta-function), $W(t)=\bar{W}(1-\cos\omega t)$ and $r^*(t)=kvt$, where k is a unit vector along the z axis. Equation (2) then reduces to

$$T=\frac{\bar{W}}{8k\kappa^{\frac{1}{2}}\pi^{\frac{1}{2}}}\int_0^t(1-\cos\omega[t-\tau])\times\exp\{-[x^2+y^2+(z-vt)^2]/4\kappa\tau\}\frac{d\tau}{\tau^{\frac{1}{2}}}.\quad(3)$$

This may be shown to give

$$T=\frac{\bar{W}}{4\pi kr}\Big\{\tfrac{1}{2}\exp[(v/2\kappa)(r+z)]\,erfc\Big(\frac{r}{2(\kappa t)^{\frac{1}{2}}}+\frac{v}{2}(t/\kappa)^{\frac{1}{2}}\Big)+\tfrac{1}{2}\exp[(-v/2\kappa)(r-z)]\,erfc\Big(\frac{r}{2(\kappa t)^{\frac{1}{2}}}-\frac{v}{2}(t/\kappa)^{\frac{1}{2}}\Big)-\exp[vz/2\kappa-rv/2\kappa(R+1/2)^{\frac{1}{2}}]\cos(\omega t-rv/2\kappa(R-1/2)^{\frac{1}{2}})-\frac{e^{vz/2\kappa}}{\pi}\int_{v^2/4\kappa^2}^{\infty}e^{-tu}\sin(r(u/\kappa-v^2/4\kappa^2)^{\frac{1}{2}})\frac{u\,du}{u^2+\omega^2}\Big\},\quad(4)$$

where $erfc(x)=\int_x^\infty\exp(-u^2)du$ and $R=(1+16\omega^2\kappa^2/v^4)^{\frac{1}{2}}$, which is seen to reduce to the expression given by Podolsky for $t\to\infty$. One may let the upper limit of integration in (3)$\to\infty$ and obtain the steady-state solution immediately if this is the portion of main interest.

The laplace transformation was used in obtaining (4) from (3). The integral in (3) is of the convolution type. A transformation with respect to time results in a simple expression which may then be inverted by contour integration.

[1] Carslaw and Jaeger, *Conduction of Heat in Solids* (Oxford University Press, London, 1947), p. 223.
[2] See reference 1, p. 216.

Books

Foundations of Aerodynamics

By A. M. Kuethe and J. D. Schetzer. Pp. 374+x. John Wiley and Sons, Inc., New York, 1950. Price $5.75.

This book is intended as an introductory text on the senior-graduate level for students of engineering, in particular of aeronautical engineering. It presupposes knowledge of undergraduate physics and mathematics through advanced calculus. Topics which are considered are the basic equations of fluid motion, including conductivity and viscosity effects. Several chapters deal with various aspects of linearized airfoil theory where viscosity plays no role. Other chapters give an introduction to turbulence, boundary layer theory, and shock-wave analysis. A useful feature of the book is the discussion of experimental facts and the limitations these impose on the applicability of theoretical results of linearized or otherwise approximate theories.

The body of theoretical aerodynamics has enormously expanded during the last twenty years. Many of the advanced tools of the mathematical physicist are needed for successful work in this field. The present book will introduce the reader to much of the modern terminology of aerodynamics and prepare the way for further reading on the subject.

Eric Reissner
Massachusetts Institute of Technology

Negative Ions (Cambridge Monographs on Physics)

By H. S. W. Massey. Pp. 136+xi. Cambridge University Press, New York, 1950. Price $2.50.

H. S. W. Massey has revised and brought up to date his 1938 tract on negative ions. This second edition is the best treatment of the subject available, as was the earlier edition when it was published.

Massey has again shown his ability to select the important points to be considered in this rather complex and confused field, and to present them in an understandable manner which is a pleasure to read.

The experimental accomplishments since 1938 with negative ions have not been sufficient to necessitate any great revisions of the methods of treatment of the basic processes. The new findings are reviewed and the appropriate modifications made in the first four chapters which deal with those processes.

The fifth and last chapter has been completely rewritten. This chapter treats the effects of negative ions in glow discharges, in the upper atomosphere, and in the atmospheres of the sun and the stars. The review of the work by Chandrasekhar and his associates on the role performed by negative ions in the atmosphere of the sun is excellent. The discussion of negative ions in the upper atmosphere of the earth is also very good.

Willard H. Bennett
University of Arkansas

Electrons and Holes in Semiconductors

By William Shockley. Pp. 558+xxiii. D. Van Nostrand Company, Inc., New York, 1951. Price $9.75.

In the words of the publisher's note on the book jacket, "This book has been prepared to provide a practical theoretical and working knowledge of the properties and uses of transistor devices for electrical engineers, physicists, designers, and students, whether or not they possess the mathematical knowledge formerly essential to an understanding of quantum theory or wave mechanics." While it is true that the nonmathematical reader can gain a sufficient understanding of transistor electronics from the descriptive material in the first part of the book for elementary design considerations, the book as a whole is not a "written-down" or "popular" presentation. The electronic behavior of semiconductors is a complex subject, and Dr. Shockley has not sacrificed accuracy for the sake of simplicity.

This treatment of the theory and practice of transistor electronics has been divided by the author into three parts: Part I, Introduction to Transistor Electronics; Part II, Descriptive Theory of Semiconductors; Part III, Quantum-Mechanical Foundations. In the 116 pages of Part I, only the simplest theoretical concepts are introduced, and the main emphasis is placed on interpretation in terms of experimental results. No knowledge of quantum mechanics is assumed. In the 187 pages of Part III is presented a fundamental development of pertinent quantum-mechanical theory, prefaced by two introductory chapters of the usual mathematical background. The presentation of quantum mechanics in Part III differs from conventional treatments in that primary emphasis is laid on calculation of numerical results. The 235 pages of Part II present a descriptive interpretation of the reasoning and results of the quantum theory, developed formally in Part III. In general throughout the book, the style of presentation is developmental rather than deductive. Dr. Shockley starts out with the simplest pictorial analogies, gradually adding refinements, restatements, and new physical concepts, until the reader is led into a final coherent picture of transistor behavior.

Dr. Shockley has made every effort to make the book palatable to readers with a wide range of background knowledge. Numerous general and specific references to the literature are included, and the book itself is internally cross-referenced. Each broad presentation of experimental or theoretical details is preceded by an outline indicating the scope and significance of material to be covered, and followed by an interpretive summary.

The mathematical symbolism used is uniform throughout the book, and a chart of all symbols used, with definitions, is included in the front of the book. In view of the emphasis on design computations an energy conversion chart has been included, and an appendix has been provided discussing the use and conversion of units in the cgs, mks, and atomic (Hartree) systems. Problems are included at the end of many of the sections.

George E. Evans
Oak Ridge National Laboratory

Analytical Absorption Spectroscopy

Edited by M. G. Mellon. Pp. 618+vii. John Wiley & Sons, Inc., New York, 1950. Price $9.00.

Although written primarily for the analytical and testing laboratory, this book merits a place on the shelves of any library even remotely concerned with analytical procedure of the type mentioned. The editor assumes responsibility for the general outline and writes one chapter of the book. Despite the fact that the book is a compilation of several authors, there is a pattern running through the individual assignments. Each presents an outline of problems, suggested solutions, a consideration of errors and sources, and a representative selection of references. The latter alone (1246 in 9 chapters) merit shelf space for any serious analyst, researcher, or student. The contributors are outstanding in their respective fields, and these reviewers feel that the individuals met the purpose stated in the preface, *viz.*, "to present, in the space available, a balanced, contemporary statement of topics he considered relevant to this respective assignment."

Having assumed the responsibility for the outline of the book and the coordination of the material, the editor has in fact striven to keep inconsistencies to a minimum. Since the analyst is not ordinarily concerned with the optics, it seems as though a chapter on the optics might have been included in a separate section, thus freeing some of the chapters of unnecessary detail. In his own chapter the editor gives us the benefit of a mass of detail which he has considered, but when dealing with data and conventions, he confesses inability to make a single justifiable recom-

mendation, even for plotting spectral absorption curves. He might have accepted the challenge and made the recommendations. Although the methods and terminology are necessarily inconsistent, some one, or some agency, must insist on clarification of terms and methods of presenting data. Having edited the several chapters and faced the confusion of some 1200 references, he was in an enviable (but assailable) position to do so. However, his own book is quite consistent, and he does show that the proper approach has been suggested by listing the terminology as endorsed by several societies.

The first chapter presents and discusses the preparation of systems for absorptimetric measurement and the inclusion of tables, and an exhaustive list of references is worthwhile to anyone interested in analytical procedure.

The third, fourth, fifth, and eighth chapters discuss the instrumentation for absorptimetry and of these the fifth chapter is the most interesting. This chapter has a complete description of several instruments and may be said to present the approach of the Bureau of Standards. It emphasizes the relation of errors in spectrophotometric measures to the purpose of the measures and gives a good resume of the causes of errors and their magnitudes.

The sixth chapter is admittedly the usual chapter on the photographic process with its inherent difficulties. The author compares photoelectric and photographic methods of photometry, evaluates both, and favors the former method.

The seventh chapter concerns itself with the applications of the ultraviolet and visual spectrophotometric data. There is hardly a phase of the study which has been neglected.

The ninth is an excellent chapter on the measurement and specification of color.

F. J. Sullivan, *St. Bonaventure University*
Mary E. Warga, *University of Pittsburgh*

Circuits in Electrical Engineering

By Charles L. Vail. Pp. 560+xv, Figs. 17.09. Prentice-Hall, Inc., New York, 1950. Price $5.75.

The field of this textbook is that of circuit analysis. The stated objective is to cover a wide variety of topics considered basic in this field and to provide a rigorous and detailed analytical treatment of the subject within a reasonable number of pages. To this end, the author feels that dc theory should be treated as a zero frequency case of ac theory, and that previously an undue allotment of study time has been given to dc theory.

The book is primarily intended as a textbook and not as a reference work. It is generously devoted to detail and problems designed to fix these elements in the mind of the student. The analytical approach has been used, but the mathematics employed are not of the difficult variety.

Since obviously two schools of thought exist, it seems to this reviewer that some disagreement will be found with the author's objections to studying dc theory before ac theory. The problem involved is pedagogical in nature; the proponents will cite logic as their argument while the opponents may be expected to offer arguments based on how the beginning student learns to best advantage. The preface of the book, however, leads one to expect greater differences in the text than is actually the case. In one regard, the proposed coalition was found more troublesome than the "traditional" approach. The laws for parameters in series and parallel combinations as treated in Chapters 6 and 7, respectively, are possible of simple derivation in the dc case, whereas the immediate introduction of both the resistance and admittance concept as employed when treating this subject from the general or ac standpoint would seem more complicated than desirable from the viewpoint of the beginner.

This is followed by Chapter 8 which deals with current, potential, and power computations, and then in Chapters 9 and 10, the author returns to laws for calculation of series and parallel combinations, respectively, when complex impedances are involved.

The emphasis on philosophy, historical background, and physical principles found in the early chapters will be useful to the beginning student. It is unfortunate that the needs of the majority have, of necessity, restricted the text from greater and continued emphasis on the general philosophy which would be more useful in the case of those students who would continue to advanced study.

Ellsworth D. Cook
General Engineering Laboratory

Superfluids, Vol. I.

By Fritz London. Pp. 161+viii. John Wiley & Sons, Inc., New York, 1950. Price $5.00.

This book is essentially a monograph which expounds the London philosophy of superconductivity. The discovery of the Meissner effect in 1934 showed that superconductivity was a true equilibrium state and led to the development of a macroscopic electrodynamics of superconductors which was consistent with thermodynamics. This macroscopic theory, first propounded by F. London and H. London in the late thirties and in recent years extended and refined by F. London himself, is the subject of the present volume.

The title is a good index of the author's approach to superconductivity. He regards the extraordinary mobility of electrons in superconductors as a superfluidity resulting from the appearance of quantum phenomena on the macroscopic plane. (The other known example of a superfluid is liquid helium II which will form the subject matter of Vol. II.) This point of view is suggestively presented in the introduction and permeates the body of the work. Professor London looks upon the macroscopic electrodynamic theory of superconductors as a framework into which a microscopic theory is eventually to be fitted. Consequently, if properly formulated, the electrodynamics should not only give a consistent phenomenological description but also suggest some of the broader features of the microscopic problem. This theme is present throughout the book and is developed at some length in the closing chapter which outlines a program for the molecular theory.

The book is not a survey of all superconducting phenomena and is confined to those aspects of the subject which can be naturally treated in terms of a macroscopic theory. Although primarily a theoretical treatment, the experiments which are pertinent to the development of the theory are for the most part adequately described.

The volume is divided into five sections. Section A treats the basic thermodynamics of superconductivity and the concept of the pure superconducting state as a true equilibrium state. Section B develops the electrodynamics of superconductors in considerable detail. The author carefully points out the distinction between his model and the "nonviscous flow" model.

Section C discusses the intermediate state in terms of a coarse-grained electrodynamics which deals with the mean values of the field and avoids the difficult problem of the detailed structure of the superconducting domains. Section D is concerned with the problem of the phase transition when the surface energy is taken into account.

The final section is devoted to formulating a "program for the molecular theory of superconductivity" suggested by some of the features of the macroscopic electrodynamics. No mention is made of the mass dependence in superconductivity nor of the lattice vibration theories of Bardeen and Frohlich, since apparently these developments occurred as the book was going to press. These topics, however, belong to a later stage in the development of superconductivity and their absence in no way detracts from the essential completeness of this book which is primarily concerned with macroscopic theories.

The book is well written. In developing each section the author starts from first principles and attempts to make the treatments as general as possible. This does not always make for easy reading.

Those working in the field of low temperature physics will welcome this work as an authoritative and stimulating treatment of the electrodynamics of superconductivity.

A minor error was noted. The formula at the foot of p. 92 apparently omits some constants.

E. MAXWELL
National Bureau of Standards

Structure of Molecules and the Chemical Bond

BY Y. K. SYRKIN AND M. E. DYATKINA. Pp. 509+xi, Figs 87. Interscience Publishers, Inc., New York, 1950. Price $8.75.

This book is based on the Russian edition of 1946. In it the authors present a clear, straightforward discussion of our knowledge of molecular structure, in essentially the same way as in books published in the United States between 1935 and 1940, but at somewhat greater length, and with inclusion of recent material.

Some important aspects of the chemical bond have been omitted. Thus, although there is a chapter on covalent bonds with partial ionic character, the electronegativity scale in relation to bond energy is not mentioned. The discussion of metals is very brief. Occasionally a subject is discussed so succinctly as hardly to be understandable; for example, values are tabulated of electron affinities of atoms as determined experimentally and theoretically, but with no statement whatever as to the nature of the experiment or of the theoretical treatment. The book presents a few novel features. The authors have proposed a new structure for the compound B_2NH_7, which has recently been verified by electron diffraction. They have introduced a novel nomenclature in their use of the term transitional structure: they say, for example, that the benzene molecule resonates between a first Kekulé structure, a second Kekulé structure, and the transitional structure, the probability distribution function for the last being the product of the wave functions for the two Kekulé structures. They have made extensive use of the concept of the transitional structure in the interpretation of data for electric dipole moments of molecules. Some errors have been introduced in the process of translation and publication in English. The most serious of these is the incorrect orientation of many of the electron distribution functions in Fig. 5. The translators have also misspelled a number of proper names, including Mulliken and Van der Waals.

LINUS PAULING
California Institute of Technology

Semi-Conductors (Methuen Monograph)

BY D. A. WRIGHT. Pp. 130+viii, Figs. 32. John Wiley & Sons, Inc., New York, 1951. Price $1.75.

The physics underlying the electrical properties of semiconductors is currently attracting widespread attention. As is usual in an expanding field, the need has arisen for an elementary account of the properties of these materials for students of the electron physics of solids. D. A. Wright has written this Methuen Monograph to meet this need. The subject is introduced by elementary considerations of the behavior of electrons in metals and the band structure of crystals. He then considers electron conductivity and emission in semiconductors, metal to semiconductor contacts, secondary emission, the oxide cathode, and some specific photoelectric cathodes.

The presentation is, on the whole, straightforward and is unencumbered by lengthy mathematical derivations. It is well supplemented by references. In places, organization is somewhat uneven. For example, the very interesting chapter on secondary emission is concerned largely with targets which are not semiconductors. In attempting to make this survey highly contemporary, the author has included some material which is admittedly speculative in nature. Recognized as such, it serves to emphasize the fluidity of the field and to delineate the degree to which it is understood.

Practical applications are indicated but not discussed, so that the thread of theory can be followed without excessive detours. In this book are the basic ideas and formulas of the physics of semiconductors, and it should adequately meet the need for which it was intended.

K. G. MCKAY
Bell Telephone Laboratories

Aerodynamics of Supersonic Flight

BY ALAN POPE. Pp. 184+vii. Pitman Publishing Corporation, New York, 1950. Price $4.00.

The author stated in the preface of the book that it was the interest of undergraduate students in "supersonics" which led him to develop an elementary one-term course of which the book was the text. Since many students at this stage have not had thermodynamics, the treatment is made "deliberately nonthermodynamic." The contents of the book are as follows: Chapter 1, Fundamental Equations; Chapter 2, Flow in a Duct; Chapter 3, Two-Dimensional Flow; Chapter 4, The Oblique Shock; Chapter 5, Approximate Theories; Chapter 6, Supersonic Wind Tunnels; Chapter 7, Three-Dimensional Flow; Appendix, Standard Atmosphere (up to 80,000 ft); Index; Answers to Problems.

The design of aircraft of conventional speed was perfected not by research fluid mechanics but by a generation of practical aerodynamic engineers who grasped the principles of incompressible fluid flow so well as to have a "feeling" of how the air "should move." Similarly the design of supersonic aircraft will not be perfected by research aerodynamicists, but will depend upon engineers who finally are able to understand the ways of supersonic flow and to feel it. To train such practicing engineers, the need of a good elementary text in supersonics is evident. However, the task of writing such a text is not easy, in fact, more difficult than writing an advanced book where all the techniques of mathematical analysis can be used. In an elementary text, the arguments must be "physical," easy to "visualize," intuitive yet correct. Help could have been had by appeal to experimental data, and by using a number of correctly drawn figures and graphs. Professor Pope has certainly attempted to do all this, and he is fairly successful. One may perhaps object to Fig. 4:10 in that it is so distorted as to give a wrong impression to the reader. Or one may object to the statement that the strong oblique shock is not observed. Or one may even object to the fact that the limitations of simple theories are not always pointed out as a warning to the reader. But such minor defects will certainly be remedied in later editions. The book's problems and answers will surely be appreciated by instructors and students alike.

H. S. TSIEN
California Institute of Technology

A Hundred Years of Physics

BY WILLIAM WILSON. Pp. 319. The Macmillan Company, New York, 1951. Price $3.50.

To write a satisfactory estimate of this little work by an eminent theoretical physicist, one should have sat in his classes at the University of London to become acquainted with the author. Time and again as I read his book, I had the wish that I might know Professor Wilson personally, to understand his particular slants on the world of natural philosophy, and to appreciate the restraint with which he usually treats topics that might excite the pen points of many writers to wilder enthusiasm. At first, I found little in this book to warm me either to the subject or to the author; but as I progressed, my opinion changed.

Considering the great number of significant achievements that have marked the century just past, it is bold to try to embrace the history of discovery and new ideas in so brief a volume. In physics, this has been the century of transition from caloric to thermodynamics, of the growth of the great generalization of conservation of energy, of kinetic theory of gases and statistical

mechanics, of the birth of quantum mechanics from Planck's studies of radiant heat, of the experimental foundation of electronic and nuclear physics. This was the century of Joule and Kelvin, of Maxwell and Boltzmann, and dozens of others. Some 380 names receive mention, as shown by the index, and it is little wonder that many subjects have had to be compressed to small compass or have been omitted entirely. But the author has somehow succeeded in selecting his subject matter with discrimination, in treating it with brevity and concise accuracy, and in moulding it into as coherent a finished product as the great variety of developments permits. Much of his material is treated on an elementary basis in a plodding, factual, and uninspired way, but the treatment shows the touch of the skilled hand. Here and there, those well acquainted with the physics of recent times can benefit by the personal turn that Professor Wilson gives to his topics. The reader can often detect the creative master explaining in his own well-considered way the work of other men.

In the short space allowed there is always competition between historic event and fact and the interpretation of ideas: the latter usually wins. As a result the reader may admire the later but yearn for more of the former. The chronological aspects of development are given secondary importance to the trend of ideas. Professor Wilson treats his topics and his personalities of physics with restraint, although here and there he breaks over into a well-deserved superlative. There is an occasional touch of humor, but generally speaking the author seems to be inhibited by habitual scientific caution and a sense of balanced judgment, for which of course the reader must respect him. Now and then, as in his treatment of relativity, he warms to his subject so much that one could wish that the whole book were written with such spirit and masterful appreciation.

On the whole, one must agree that William Wilson has succeeded in attaining the aim expressed in his preface where he states: "The discovery of the cardinal facts of physics and the growth of its unifying theoretical structure are described together. There is much in it that is addressed to both students and teachers of science and a little of it, here and there, even to the great pundits of physics." As many of those who read this review will naturally class themselves among the pundits (!), they may be happily surprised in searching through Professor Wilson's little volume. Like the Princess of Serendip, they may find valuable nuggets of concise and penetrating insight scattered throughout this modest treatment of an enormous body of human knowledge and its recent development. The metaphor is not inept, for gold nuggets and diamonds are often to be found by the trained eye in the midst of masses of sand and gravel of lesser value.

RICHARD M. SUTTON
Haverford College and Case Institute

Electric Illumination

Second edition. By JOHN O. KRAEHENBUEHL. Pp. 446+ix. John Wiley & Sons, Inc., New York, 1951. Price $8.00.

This is the second edition of a book published some ten years ago. As stated in the preface it deals with the principles underlying the specification and design of electrical lighting for commercial and industrial buildings, and is intended to cover the needs of architects, and of architectural and electrical engineers.

There are 13 chapters in the book that deal with various phases of the lighting field, such as specifications of illumination, seeing, light sources, control of light, and also chapters on flood lighting and novelty lighting, since such information is often in demand. The book is well illustrated with 148 figures and 71 tables.

The nomenclature of Illuminating Engineering, as laid down by the special committee of that society, is given with good explanations and illustrations. A consideration of the amount and quality of the illumination, for various seeing tasks, is presented, again with well-chosen illustrations. The methods of calculating the illumination given by various sources are outlined in Chapter 5, and Chapter 6 gives a good short description of the construction and operating conditions of various electric lamps. Various methods of light control are given the prominent place they warrant.

The important subject of the maintenance of a lighting system comes in for some apt remarks. A good point is made of the fact that today one can get the higher illuminations recommended at less cost than was possible some years ago, due to the higher efficiency of the lamps and the lower cost of both the lamps and the operating electric energy. The final chapter on wiring is a good addition since this is a subject that laymen—and indeed some scientists—are not too familiar with.

Some errors have crept into the text, e.g., (pp. 18 and 28) the inverse square law and the size of the source; the ordinate of Fig. 6–4; candlepower is not a unit (pp. 24 and 113); many will not agree that gaseous discharge lamps have only one advantage (p. 155). Such terms as visible light and ultraviolet light should be avoided (p. 363). Today the hum due to the choke used with the fluorescent lamp is practically eliminated (p. 166). It seems too bad, considering the advance in this subject, that the table of recommended illuminations for various tasks should be taken from a report that is ten years old (p. 167). Also that the problem on the economics of the fluorescent lamp should quote a paper that was published more that ten years ago when this lamp was so new (p. 390). For some it would be better if the references were all marked in the text so that it would be easier to find a particular reference.

Taken all in all this is a very readable book and one that a teacher should find helpful in presenting the subject of light and lighting.

W. E. FORSYTHE
General Electric Company

Basic Wing and Airfoil Theory

By ALAN POPE. Pp. 294+xviii. McGraw-Hill Book Company, Inc., New York, 1951. Price. $5.00.

As the author points out, this book is a text for college senior or first-year graduate students. Following two short and clearly written chapters treating vectors, complex variables, and other mathematical aids necessary for the understanding of the text, a more or less conventional introduction is given to stream and potential functions. Source-sink and vortex flows are treated and applied specifically to the case of the circular cylinder, and Kutta-Joukowski theorem of lift is derived. The next four chapters concern aerodynamics of airfoil sections. Derivation of airfoil sections by Joukowski transformation is given with suitable examples. Methods of thin airfoil theory are introduced, and a means for separating profile thickness and chamber effects is shown. The theory of thick airfoils of arbitrary shape or pressure distribution with the method for estimation of airfoil drag completes the treatment of sections. The next four chapters treat planform characteristics. Following a conventional treatment of the Biot-Savart law and trailing vortices, downwash is considered, and the concept of induced drag is given. Effects of aspect ratio and planform shape are treated. Methods of calculation of downwash and sidewash are given, followed by a short treatment of effects of sweepback and dihedral. Then span-load distribution for arbitrary wings with determination of stall pattern is provided. The final chapter of the book treats miscellaneous applications of perfect fluid theory (e.g., wind-tunnel wall interference). This well-written and well-balanced text is an improvement over older texts, primarily in that the more modern methods for treatment of airfoil sections (particularly thin airfoil methods) and for calculation of span-load distributions (Multhopp, Schrenk, etc.) are given. Example problems are well chosen.

H. JULIAN ALLEN
Ames Aeronautical Laboratory

Other Books Received

Elements of Television Systems. GEORGE E. ANNER. Pp. 803+viii. Prentice-Hall, Inc., New York, 1951. Price $7.75.

An International Bibliography on Atomic Energy, Vol. 2. ATOMIC ENERGY COMMISSION GROUP, Department of Security Council Affairs, United Nations. The Columbia University Press, New York, 1951. Price $10.00.

Galvanizing (Hot-Dip). Third Edition. HEINZ BABLIK. Pp. 502, Figs. 306. John Wiley and Sons, Inc., New York, 1951. Price $10.00.

Grundlagen der Atomphysik. HANS ADOLF BAUER. Pp. 631 +xx, Figs. 244. Springer-Verlag, Wien, 1951.

Applied Electronics Annual 1951. EDITED BY R. B. BLAISE. Pp. 263. J. D. Griffiths, 2 Clarke Court, Rutherford, New Jersey, 1951. Price $8.00.

Physik und Technik der Ultrarotstrahlung. WERNER BRUGEL. Pp. 257. Curt R. Vincentz Verlag, Hannover, 1951.

Tables of the Bessel Functions of the First Kind of Orders Seventy-Nine Through One Hundred Thirty-Five. STAFF OF THE COMPUTATION LABORATORY, Harvard University. Pp. 614. Harvard University Press, Cambridge, Massachusetts, 1951. Price $8.00.

Tables Relating to Mathieu Functions. THE COMPUTATION LABORATORY OF THE NATIONAL BUREAU OF STANDARDS. Pp. 278+xvii. Columbia University Press, New York, 1951. Price $8.00.

Synchronous Machines. CHARLES CONCORDIA. Pp. 224, Figs. 23. John Wiley and Sons, Inc., New York, 1951. Price $5.50.

Industrial Instrumentation. DONALD P. ECKMAN. Pp. 396 +vi. John Wiley and Sons, Inc., New York, 1950. Price $5.00.

Radiation Monitoring in Atomic Defense. DWIGHT E. GRAY AND JOHN H. MARTENS. Pp. 122+vi, Figs. 20. D. Van Nostrand Company, Inc., New York, 1951. Price $2.00.

Thermodynamics, Second Edition. GEORGE A. HAWKINS. Pp. 563+xv. John Wiley and Sons, Inc., New York, 1951. Price $6.50.

Analysis of the Four Bar Linkage. J. A. HRONES AND G. L. NELSON. Pp. 730+xx, Figs. 730. The Technology Press of the Massachusetts Institute of Technology and John Wiley and Sons, Inc., New York, 1951. Price $15.00.

Short Wave Wireless Communication, Fifth Edition. A. W. LADNER AND C. R. STONER. Pp. 717+xvi, Figs. 417. John Wiley and Sons, Inc., New York, 1951, Price $8.00.

Quantum Mechanics. ALFRED LANDÉ. Pp. 307+x. Pitman Publishing Corporation, New York, 1951. Price $5.50.

Elements of Modern Physics. WALTER C. MICHELS AND A. L. PATTERSON. Pp. 659+ix. D. Van Nostrand Company, Inc., New York, 1951. Price $6.50.

Jacobian Elliptic Function Tables. L. M. MILNE-THOMSON. Pp. 132+xi. Dover Publications, Inc., New York, 1951. Price $2.45.

Jacobian Elliptic Functions. ERIC HAROLD NEVILLE. Pp. 345+xiv, Figs. 32. Oxford University Press, New York, 1951. Price $7.00.

Civil Defense in Modern War. AUGUSTIN M. PRENTISS. Pp. 429+x, Figs. 60. McGraw-Hill Book Company, Inc., New York, 1951. Price $6.00.

Workbook for Weather Forecasting. E. S. PULK AND E. A. MURPHY. Pp. 66+vi, and map series kit. Prentice-Hall, Inc., New York, 1951. Price $3.75.

Fundamentals of Electrical Engineering. FRED H. PUMPHREY. Pp. 668+xii. Prentice-Hall, Inc., New York, 1951. Price $5.75.

The Elements of Mathematical Logic. PAUL ROSENBLOOM. Pp. 214+iv. Dover Publications, Inc., New York, 1951. Price $2.95.

The Atom at Work. JACOB SACKS. Pp. 327+xii, Figs. 54. The Ronald Press Company, New York, 1951. Price $4.00.

Fundamentals of Physics. HENRY SEMAT. Pp. 849+xxv. Rinehart and Company, Inc., New York, 1951. Price $6.00.

Reports on Progress in Physics, Vol. XIV. EDITED BY A. C. STICKLAND. Pp. 412. The Physical Society, London, 1951.

Borderland of Science. ALFRED STILL. Pp. 424+ix. The Philosophical Library, Inc., New York, 1951. Price $3.75.

Elements of Ore Dressing. ARTHUR F. TAGGART. Pp. 505 +xvii. John Wiley and Sons, Inc., New York, 1951. $10.00.

Space Time Matter. HERMANN WEYL. Pp. 330+xiv, Figs. 15. Dover Publications, Inc., New York, 1951. Price $3.95.

The Fourier Integral. NORBERT WIENER. Pp. 201+ . Dover Publications, Inc., New York, 1951. Price $3.95.

The Preparation of Programs for an Electronic Digital Computer. MAURICE V. WILKES, DAVID J. WHEELER, AND STANLEY GILL. Pp. 167. Addison-Wesley Press, Inc., Cambridge, Massachusetts, 1951, Price $5.00.